An Evaluation of International Surveys of Children

VS/2010/0245 (SI2.561486)

December, 2012

Dominic Richardson

Social Policy Division

www.oecd.org/els/social

Directorate for Employment, Labour and Social Affairs

TABLE OF CONTENTS

Boxes

Figures

Tables

AN EVALUATION OF INTERNATIONAL SURVEYS OF CHILDREN

This publication is supported by the European Union Programme for Employment and Social Solidarity - PROGRESS (2007-2013).

This programme is implemented by the European Commission. It was established to financially support the implementation of the objectives of the European Union in the employment, social affairs and equal opportunities area, and thereby contribute to the achievement of the Europe 2020 Strategy goals in these fields.

The seven-year Programme targets all stakeholders who can help shape the development of appropriate and effective employment and social legislation and policies, across the EU-27, EFTA-EEA and EU candidate and pre-candidate countries.

For more information see: http://ec.europa.eu/progress

The information contained in this publication does not necessarily reflect the position or opinion of the European Commission.

Acknoweldgements

Support for the statistical work was given by Nabil Ali and Sophie Bournot (both statisticians in the Social Policy Division at the OECD).

Thanks go to the survey coordinators who have given their time to review and comment on the work, and have, when needed, facilitated access to information and data included in the analyses.

General disclaimer

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities, the use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Important note: This report has been written over the course of two years, and every effort has been made to ensure that information is accurate and up-to-date. However, the nature of the surveys means they are regularly repeated and constantly updated in terms of content and methods, and as such readers should refer to survey-specific studies, websites and technical reports for the most up-to-date data and information.

EXECUTIVE SUMMARY OF THE KEY QUESTIONS AND MAIN FINDINGS

In January 2008, the European Commission adopted a report on child poverty and well-being (EC, 2008), including 15 recommendations to enhance the monitoring of child poverty and child well-being at EU and national levels. Recommendation 12 states that "an in-depth evaluation of the available international data sources covering important aspects of child well-being is needed in order to assess whether these sources can supplement existing national and/or transnational sources". The purpose of this report is to directly address this recommendation, and assess the quality and availability of international data for child well-being measurement to supplement national and international sources. This executive summary introduces the objectives and structure of the report, setting out the key questions to be addressed. It concludes with a summary of the main findings of the work.

i. The objectives and structure of the report

1. The objective of this project is to evaluate sources of international child well-being data to assess their suitability for supplementing national and transnational sources – as outlined in recommendation 12 of the report *Child poverty and well-being in the EU: Current Status and the Way Forward* (EC, 2008). The evaluation will inform cross-national monitoring of children's lives, identify gaps in measuring child well-being not covered by available international data sources, and provide recommendations for the use and improvement of international sources of data used for the monitoring of child well-being. Three research questions will be addressed:

- What available data can be used to assess the well-being of children across countries?
- What are the strengths and weaknesses of existing international surveys for informing policy and monitoring the lives of children?
- How can the data for cross-national monitoring of child well-being be improved?

2. To answer these three questions, four activities will be undertaken (outlined in detail in section 1.4):

- A systematic review of the international data sources on children outlining the content of the survey datasets, methodology, and organisational processes.
- An evaluation of the strengths and weaknesses of international data sources based on empirical testing of common indicators of child well-being derived from the various surveys.
- A review of the gaps in the child well-being data availability, organised by child characteristics and public policy dimensions, and suggestions for ways in which the data for cross-national monitoring of child well-being can be supplemented and improved upon.
- A presentation of a set of recommendations for the use and improvement of data derived from international surveys of children, including how gaps in the data might be filled and provide methodological recommendations for the use of international child datasets for monitoring of child well-being across countries (see 2.4.4).

3. Each activity will be addressed in separate chapters, following on from Chapter 1, which introduces the motivation behind the report, the selection criteria for surveys included in the study, and the

main questions to be answered in the four upcoming chapters. Chapter 2 systematically reviews the surveys' content, methods and organisational structure. Chapter 3 empirically evaluates the surveys for validity and reliability across key child well-being indicators. Chapter 4 organises and reviews the content of the surveys to identify gaps in the coverage of child ages and the content suitable for child well-being comparisons. And Chapter 5 brings together the findings to provide recommendations on the use and improvement of international surveys of children.

ii. Key findings and recommendations

4. Drawing from the review of the survey methodology, a stock take of the available indicators, and an evaluation of the most commonly used child well-being items from those surveys; the report concludes with a set of key recommendations. The following list summarises the main recommendations to be addressed in the use and improvement of available data for child well-being policy, research and monitoring:

1. **Concerted efforts should be made in policy and research circles to fill gaps in child well-being comparisons** in terms of both age-related indicators (children under nine are missing from survey work) and in terms of new dimensions and indicators of child well-being, not presently covered in the studies (child protection, mental health measures, and more recently civic participation). There is a role for all parties in this process; from the demand-side in terms of policymakers and researchers, and the supply-side from funders and survey coordinators. Although some series data is available, survey data is needed for detailed analysis and recommendations. Some survey data is already available to do this in Europe (ESS and ESPAD), for non-European countries new data (possibly via amendments to present collections) are needed.

2. **Communication between survey coordinators and survey users can be improved.** For instance, surveys should be encouraged to communicate their proposed changes wave-on-wave, and consult users (through open forums or bilateral discussions). Policymakers and researchers should be encouraged to provide evidence in support of claims, with respect to the stated purpose and goal of the survey in question. This will help with planning for monitoring and research purpose, and enable survey coordinators to reflect on priorities. These communications could be built on examples of ongoing dialogue between survey coordinators themselves (see section 5.4.1).

3. **The use of equality indicators and social gradient indicators in monitoring child well-being are necessary.** Policymakers and researchers need to use equity indicators when assessing outcomes from public services. Researchers should use available data to explore inequality across more indicators, and surveys should be encouraged to produce scalable items in new questionnaires and / or amendments to present items.

4. **All new child well-being indicators derived from secondary analysis of surveys should be validated, and if necessary, treated for non-response bias.** Researchers should undertake post-hoc weighting adjustments to indictors with bias in non-response (see 5.5.2); when more complex multivariate analysis is undertaken, to ensure representative estimates to fill gaps are needed. Policymakers should ensure that validation tests and appropriate adjustments have been undertaken before allowing the findings of secondary analysis to directly inform policy decisions. Survey coordinators should provide guidance, when relevant to their specific survey data, to cover methods for validation or treatment of non-response bias in their surveys.

5. **All indicators and analysis derived from present school/household surveys should clearly show which children are missing from the analysis, and effort should be made to fill those gaps** (see 5.5.1). To facilitate this, in the first instance, survey coordinators should report on missing populations (with the help national survey managers), or consider supplemental out-of school surveys.

6. **Efforts should be made to harmonise data collections between the major international surveys**, but a balance should be struck to protect the unique contribution of each survey. This will facilitate analysis on interactions between various well-being experiences and provide new evidence for policy presently not available. Where national sets have been quality checked, national efforts to harmonise international evidence should also be undertaken. Funding bodies or governmental bodies that work with more than one survey can be involved in these efforts. Examples of how to do this exist, HBSC and ESPAD have already harmonised on some items of substance use.

7. **Notwithstanding financial constraints, funders and policymakers should be prepared to support a more comprehensive international survey of children** (by child well-being content and age) if present surveys cannot adapt to meet the necessary requirements for properly informing policies that aspire to cross-sectoral integrated interventions for *all children*. Any survey should strive to innovate in this area, and two key ways of doing this include a longitudinal aspect to the approach, an attempt to capture all children, and a commitment to a multidimensional approach.

8. **If efforts to improve the knowledge base for comparing child well-being are to be meaningful, it is essential that policymakers respond to the availability of new data and evidence with advances in monitoring and policy development.** In doing this, it is necessary to consider which indicators are most appropriate for informing policies for childhood in different age-groups, and in different at-risk categories, and use them in target-setting and the retrospective evaluation of both new *and* on-going policies for children.

CHAPTER 1. REVIEWING INTERNATIONAL DATA ON CHILDREN'S WELL-BEING: MOTIVATION AND METHODS

This chapter introduces the motivations and the methods for the project, looking at the review, evaluation, gaps, and recommendations sections in turn. It provides the rationale for the focussing on specific surveys for review, and in terms of limiting the overall analysis. In recent years child well-being comparisons have served to show two important things: the first is that there is great interest – politically and publically – in understanding what contributes to a good childhood, and how this should be effectively monitored cross-nationally. And second, that the data that is available to researchers and policymakers – for the purposes of monitoring child well-being holistically – is incomplete in terms of age coverage and content, and lacks the confidence given to adult surveys used in decision making. The chapter gives important details to the analytical frame of the project, and in doing so sets out the scope of the work.

1.1. Introduction

5. To assess whether existing surveys of children, and households with children, can supplement national or transnational sources used for monitoring living standards in advanced economies some basic questions need to be addressed. Why are child well-being comparisons necessary, and why is an evaluation of available data needed? How many countries do the surveys cover, are they regular surveys with future waves planned, and do they collect data from children (0-17 years of age)? Does the content match presently understood concepts of child well-being measurement, and how might the content quality be assessed?

6. To answer these questions, this chapter provides the motivational background to this work, the selection criteria for the surveys to be reviewed, as well as an outline of the project's analytical framework.

7. The first section of this chapter introduces the motivation for this project. The second section identifies surveys to be reviewed, the countries that they cover, and the selection criteria for inclusion in the project. The third section presents a detailed analytical framework for each of the following chapters, including: the process for the systematic review of surveys, the methods of evaluation of the selected surveys, the coverage of the available data and indicators in those sets, and the delivery recommendations for the use and improvement of international survey-derived data on children and childhood.

1.2. Motivation: Why are child well-being comparisons necessary, and why is an evaluation needed?

8. Child well-being is an increasing concern for policymakers. How children fare through critical points of development will affect their quality of life, their productivity, welfare dependency and the transmission of their later life outcomes to their own children. In response to this increasing concern, recent years have seen a growth in the international comparisons of child well-being (Bradshaw et al, 2007; Bradshaw and Richardson, 2009; EC, 2008; UNICEF IRC, 2007; and OECD 2009 and 2011): comparisons that have had both a political and public impact.

9. Alongside this growth in child well-being comparisons has been a growing political interest to move beyond the sole use of income poverty measures to assess children's life outcomes over time (TARKI/APPLICA, 2010). Child income poverty has been a headline indicator for children's living

standards for over a generation of children, but it has come under scrutiny for not capturing comprehensive life experiences, being calculated in an arbitrary fashion (both equivalisation and thresholds used), for ignoring the detailed needs of children by age or disability, and ignoring credit, saving, gifts and home production. Recent highly publicised cross-national comparisons of child well-being have shown the range of indicators available for monitoring children's living standards, and as such led to calls for greater openness and a clearer understanding of the data underlying these projects.

10. Interest in comparing child well-being and evaluating the robustness of the comparisons is in the interests of the public and policymakers alike, who to know how the reality of how children, in the most economically advanced countries, are faring.

11. Moreover, the evidence to come out of these recent cross-national comparisons of child well-being indicators have added to need to better understand the quality of the data, and what the data truly represents. First, recent reports have shown that well-being measures can evolve in different directions (such as infant mortality and low birth weights – see OECD, 2011). Second, that virtuous tradeoffs between efficiency and equity outcomes are evident (Freeman *et al*, 2010). Third, poverty, family affluence and family structure, do not associate strongly to many important child-well-being indicators. Fourth, that although public spending is associated to child well-being measures, some countries are getting more for their money (OECD 2009, 2011; Förster and Richardson, 2011). And finally, that some indicators are more malleable over the short term than others (for instance child reading literacy rates in comparison to child poverty rates).

12. For these findings to feed into policy decisions, the need for confidence in the reported child well-being statistics cannot be underestimated. Data on children, and the quality of their lives, will feed in to public debate about social expenditure in the context of several other important competing demands. If figures are incorrect or misguiding, limited public money might be wasted to the detriment of the competing demand on the public budget, and societal development on the whole. Moreover misrepresenting the actual state of children lives across countries, or within countries on the basis of socio-economic or socio-demographic differences, can be counterproductive for children themselves insofar as targeted resources may a) miss the children they are designed to help, and b) erode public confidence in similar evidence base responses in the future, to the detriment of all children.

1.3. Surveys to be reviewed: questions of when, who, how, where?

13. Six child surveys will be reviewed as part of this project, including:

- the European School Project on Alcohol and other Drugs (ESPAD),
- the Health Behaviour in School-aged Children (HBSC) study,
- the International Civic and Citizenship Education Study (ICCS),
- the Progress in International Reading Literacy Study (PIRLS),
- the Programme for International Student Assessment (PISA), and
- the Trends in International Mathematics and Science Study (TIMSS).

14. Table 1 outlines the surveys and country membership in the most recent and upcoming surveys. Across the OECD Japan and Mexico take part in only two of the six surveys reviewed. All but two European Union (EU 27) countries part in at least four of the six surveys each (Cyprus[1,2] and Portugal).

[1] Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognizes the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable

Seventeen countries in total are members of all six surveys, although on occasion Belgium is represented by one community, and the United Kingdom by at least one home country (England, Northern Ireland, Scotland or Wales). Each of these 17 countries is a member of the EU 27. With the exception of the Bulgaria, Latvia and Lithuania these countries are also members of the OECD.

Table 1: Countries participating in the six selected child surveys

	EU School Project on Alcohol and other Drugs (ESPAD)	Health Behaviour in School-aged Children Study (HBSC)	International Citizen and civic survey (ICCS)	Progress in International Reading Literacy Study (PIRLS)	Programme for International Student Assessment (PISA)	Trends in International Mathematics Science Study (TIMSS)
Australia	x	x	x	✓	✓	✓
Austria	✓	✓	✓	✓	✓	✓
Belgium	✓	✓	✓	✓	✓	✓
Bulgaria	✓	✓	✓	✓	✓	✓
Canada	x	✓	x	✓	✓	✓
Chile	x	x	✓	x	✓	✓
Cyprus[1,2]	✓	x	✓	x	x	✓
Czech Republic	✓	✓	✓	✓	✓	✓
Denmark	✓	✓	✓	✓	✓	✓
Estonia	✓	✓	✓	x	✓	x
Finland	✓	✓	✓	✓	✓	✓
France	✓	✓	x	✓	✓	x
Germany	✓	✓	x	✓	✓	✓
Greece	✓	✓	✓	x	✓	x
Hungary	✓	✓	x	✓	✓	✓
Iceland	✓	✓	x	✓	✓	x
Ireland	✓	✓	✓	✓	✓	✓
Israel[1]	x	✓	x	✓	✓	✓
Italy	✓	✓	✓	✓	✓	✓
Japan	x	x	x	x	✓	✓
Korea	x	x	✓	x	✓	✓
Latvia	✓	✓	✓	✓	✓	✓
Lithuania	✓	✓	✓	✓	✓	✓
Luxembourg	x	✓	✓	✓	✓	x
Malta	✓	✓	✓	✓	x	✓
Mexico	x	x	✓	x	✓	x
Netherlands	✓	✓	✓	✓	✓	✓
New Zealand	x	x	✓	✓	✓	✓
Norway	✓	✓	✓	✓	✓	✓
Poland	✓	✓	✓	✓	✓	✓
Portugal	✓	✓	x	x	✓	x
Romania	✓	✓	x	✓	✓	✓
Slovak Rep	✓	✓	✓	✓	✓	✓
Slovenia	✓	✓	✓	✓	✓	✓
Spain	x	✓	✓	✓	✓	✓
Sweden	✓	✓	✓	✓	✓	✓
Switzerland	✓	✓	✓	x	✓	✓
Turkey	✓	✓	x	x	✓	✓
United Kingdom	✓	✓	✓	✓	✓	✓
United States	x	✓	x	✓	✓	✓

Notes: Results are based on available evidence at the time of writing in 2011 – updated lists for some surveys are available in Annex 2. Shaded cells highlight countries not covered by the surveys. In HBSC Belgium is sampled in Flemish and French Belgian communities separately, the UK is sampled as England, Scotland and Wales separately. The provinces of Alberta, Ontario, and Quebec are Canadian benchmarking participants in 2011 for both TIMSS and PIRLS. In PIRLS 2011 the UK is represented by English and Scottish samples only. PIRLS 2006 countries not in 2011 include Flemish Belgium, Iceland, and Latvia. TIMSS 2007 countries not in 2011 include: Bulgaria, French Belgium, Cyprus, and Latvia. In ESPAD 2007 Germany is represented by 7 Lander, Turkey sampled 6 cities in 2003 but did not take part in ESPAD 2007. French Belgium did not take part in ICCS 2009 and only England took part from the United Kingdom.

1 - The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Sources: Survey websites, Accessed May 2010.

---

solution is found within the context of United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

2 Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognized by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

15. Three household panel surveys are included in the evaluation: the European Survey of Income and Living Conditions (EU SILC) the European Quality of Life Survey (EQLS) and the European Social Survey (ESS). These surveys are included, not simply because they ask about the numbers of children within any given household, but because data collected by the surveys are specifically designed to generate findings about the living conditions of children, or families with children. Because each of these surveys is a European household survey they do not include any non-European OECD members countries (with the exception of Turkey in EQLS), and they do not ask questions directly to dependent children under 18 (but head of households from age 15; from age 16 for EU-SILC).

16. A further survey, the International Study of Computer and Information Literacy, will be undertaken in 2013. Because data will not be available until 2014 or beyond, the survey is not included in the evaluation (see http://www.iea.nl/icils.html for more details).

17. The information provided by five of these surveys has been used in a range of international comparisons of child well-being. ESPAD has been used in the first comparison of child well being in the EU by Bradshaw and colleagues (2007), PISA and HBSC surveys have been ever present in European (*ibid*, and Bradshaw and Richardson, 2009) and OECD comparisons of child well-being (UNICEF, 2007 and OECD, 2009). Both TIMSS and PIRLS are included in the this project because they meet the selection criteria of this evaluation, but neither has had sufficient numbers of EU or OECD countries in the past to be included in the cross-national comparisons listed above, which tend to use more strict selection criteria for the inclusion of data sources (75% of countries covered, not 50% as this study). ICCS is a new survey which reported for the first time in June 2010 (ICCS, 2010) and so it has not yet contributed to any of the main child well-being indicator frameworks.

1.3.1 Survey selection process

18. The child surveys included in this project meet four selection criteria (when, who, how and where):

1. **The survey must be a current survey with plans for future waves** – reviewing surveys that will not be repeated defeats the object of the exercise. Surveys must be able to supplement existing international and national sources of child data, both today and in the future.
2. **The survey must ask questions directly to, or about, children under the age of 18 (the UNCRC definition of childhood)** – because improving the knowledge base for child well-being monitoring is a key objective of the report, it is of critical importance that evidence is taken from surveys that record the experiences of children during childhood.
3. **The content of the survey must reflect in some way an aspect of child well-being that is unique to that survey** – bearing in mind the limited data situation already faced by analysts in this area, this project seeks to maximise the coverage of unique content over different age groups of children.
4. **At least half of the OECD countries (17 of 34) or European countries (14 of 27) must be included in the survey** – particularly in light of the objective to supplement existing international sources, country coverage is a key factor in inclusion. Surveys with insufficient numbers of EU or OECD countries to be included in international multi-dimensional child well-being reports are not reviewed simply because they would be excluded from monitoring plans on this criterion alone. The low threshold (50%) for the selection of surveys into the review is because of the potential surveys have for expansion.

19. Each of the child surveys listed in Table 1 have met those criteria. Table 2 below shows the coverage of children in each survey by age, sample type, the unique content of the survey.

Table 2: Criteria for child survey selection

	EU School Project on Alcohol and other Drugs (ESPAD)	Health Behaviour in School-aged Children Study (HBSC)	International Civic and Citizenship Education Study (ICCS)	Progress in International Reading Literacy Study (PIRLS)	Programme for International Student Assessment (PISA)	Trends in International Mathematics Science Study (TIMSS)
Most recent wave (next wave)	**2007** (2011, 2015)	**2009/10** (20013/14)	**2009**	**2006** (2011, 2016)	**2009** (2012, 2015)	**2007** (2011, 2015)
Approx. Child age	16 year olds	11, 13 and 15 year olds	8th grade students (above the average age of 13.5 years)	4th grade students (aged around 9.5 years)	15 year olds	4th and 8th grade students (aged around 9.5 and 13.5 years)
Content of focus	Alcohol, tobacco and drug use	Health status and behaviours	Civil societies and systems, civic principles, civic participation, and civic identities	Reading literacy (curricula based)	Reading, mathematics and science literacy (life skills-based measure)	Mathematics and Science abilities (curricula based)
Countries (EU/OECD)	25/23	26/28	22/23	23/25	25/34	22/27

Source: OECD.

Note: Information in the table, and following text, refers to surveys undertaken in the years in bold.

20. The surveys selected run in waves of three (PISA) four (ESPAD, HBSC, TIMSS) or five (PIRLS) years – most recently undertaken in 2009, 2010 or 2011.[3] In each of the surveys at least 23 OECD countries are included and at least 22 EU 27 countries. The six child surveys collect data from children between the ages of around 9 and 16 years. Four of the surveys primarily focus on aspects of education achievement in mathematics, reading, science and citizenship studies (ICCS, PIRLS, PISA and TIMSS) and two on aspects of health and health behaviours (ESPAD and HBSC). Because each of the surveys focuses on different well-being data, not all of these aspects of well-being are available for children of all ages.

1.4 How to evaluate the content of international surveys of children? An outline of the analytical framework

21. The following sections briefly outline the methods for reviewing, evaluating, checking the content of, and developing recommendations for the use and improvement of international surveys of children.

1.4.1 Systematic review of international datasets on children

22. A key starting point for assessing the availability and quality of comparative data of children is to review the surveys from which the data is derived. Chapter 2 undertakes a systematic review of the six child surveys and three household surveys. Each part of the systematic review contributes to the work undertaken in the following Chapters of the project. The survey information is organised into one of the following themes: statement of purpose, management process and funding; basic dataset information

[3] Surveys waves with available data are included in the review in order to match the coverage with the evaluation of data undertaken in Chapter 4.

(including content); questionnaire development and sampling strategies; and methods of field collection. The final section of the review provides details about the future waves of the surveys.

23. To populate the sections of the review, background information and technical reports for each survey has been consulted.

24. Chapter 2 discusses the results of the review with a particular focus on how the choices made for each survey will impact on its ability to contribute to comparisons, or monitoring, of child well-being.

1.4.2. Evaluation of the results of international surveys of children

25. International surveys of children seek to identify the same 'target' population of children across countries to answer the same set of questions, whilst ensuring comparability and reducing system-based, methodological and cultural bias in the responses. Chapter 3 evaluates how well this aim is met using varied sampling strategies, questionnaire development techniques, and other processes identified in Chapter 2.

26. The structure of Chapter 3 is organised by biases that are system-based, sample or methodologically based, and cultural or linguistically based. The empirical evaluation of the comparability and validity of the indicators in the surveys will review examples in the literature of efforts to explore forms of bias in the surveys outlined above before going on to perform tests on the data itself. Chapter 3 concludes by identifying methods which have the greatest success in reducing biases of these types. The surveys which apply these methods best will be identified for 'good practise' purposes and highlighted in the recommendations in Chapter 5. Where biases are indentified, implications for appropriate interpretation of the survey results will be made.

27. The three sections of system-based, methodological and cultural bias are introduced below with a rationale for selection and plans for analysis. In each evaluation section, evidence of bias that varies significantly between countries is raised as and when relevant. Although the data used in simple cross-national reports and monitoring inferential or descriptive statistics may be affected if raw reported values are biased, more complicated statistical analysis might only be affected by bias if this significantly varies from country to country impacting on the standardised (variance driven) results produced by multivariate models (Hanuschek and Woessmann, 2010, pp 9-10). Efforts to differentiate between these forms of bias will be made in all parts of the analysis.

1.4.2.1 System-based bias

28. The first section of Chapter 3 addresses system-based bias which refers to bias arising from the organisation of school systems, where many of the surveys are undertaken. Where the organisation of country education, and child policy systems, differ there is potential for the survey process to bias the responses of some groups of children, or countries, and in doing so affect comparability. Where interventions by governments can affect children's responses to surveys, and not their capacity to respond, they are not treated as drivers of system-based bias (curricula, school independence and so on). For this reason system-based biases refer only to the variation in organisation and timing of services for children, and issues of population size and enrolment, that are common across countries. To explore whether system-based bias exists, in education interventions for children including the timing of the school year and examinations, children enrolled in mainstream schools, and population size will all be tested.

29. To test for system-based bias access to data on enrolment rates in schools for the test cohort and figures on children 'out-of-school' as a group are needed. Chapter 3 project re-runs descriptive analysis undertaken in research by Hanuschek and Woessmann (2010), adding where possible information on dropout rates from school by sex.

1.4.2.2 Sampling and methodological bias

30. The second section of the evaluation addresses sampling and methodological bias, which refers to bias arising from the selection of sampling methods, response rates for schools and children, and questionnaire methodology (such as structure and order of items).

31. For example, child populations vary massively across OECD countries, and so sample strategies which rely on absolute numbers of children responding (a sample of 1000 whether the country is Iceland or the US) will impact on the extent of statistical reliance on applied weighting methods, on techniques applied to infer confidence in population estimates, and on the comparability in sub-populations. In contrast, proportional samples which apply response rate thresholds to samples of schools and pupils, to ensure data is comparable across countries, can ignore the fact that non-respondents may be significantly different from respondents and that non-response might be non-random.

32. Evidence in the literature on non-response in England's PISA results (Micklewright *et al*, 2010) has shown clear evidence that upward bias exists in both the 2000 and 2003 samples for the mean score even though 2000 data met the international inclusion criteria and 2003 data were rejected due to low response rates (below arbitrary response rate thresholds of 80% of schools and pupils after replacement sampling). The authors also find evidence of downward bias in the standard deviations in national mean results. However as the Micklewright *et al* point out (*ibid*) it is often difficult to test for non-response bias due to the lack of data available for individuals who do not respond to the survey. To overcome these problems researchers can data match to administrative data, or test for non-response bias on specific items within the survey itself where non-respondent data is available for other items.

33. Questionnaire methodology may also impact on the quality and usability of results. For example placement of complex items in a questionnaire (early or late) may affect the quality of the responses received, insofar as children's concentration levels may fall, to the point where some children may not even answer the question.

34. Chapter 3 will test selected items from the child and households surveys for associations with population sizes, sample sizes and participation rates. The questionnaire design will also be tested for evidence of how changes to the order or wording of selected well-being items affect response rates.

1.4.2.3 Cultural and linguistic bias

35. The third section of the evaluation addresses cultural and linguistic bias, which refers to bias that appears when national attributes and values affect the ability of children in that country to respond to the survey as children in other countries might. Differences in translation methods and choices regarding wording, examples or items included in a survey are areas where such biases can occur. In order to be confident that the results of an international assessment are truly comparable there needs to be confidence that the translation of words, sentences and sentiments within the questions and answers are equivalent. Furthermore, there is the potential for assessment questions to be subject to cultural interpretation that renders them incomparable – the same may be said for background questionnaires.

36. Evidence in the literature has evidenced cultural bias in survey data (Blum et al, 2001), and challenged the premise that items are comparable across cultures and languages after translation and re-translation due to use of national or regional terms or experiences. The accuracy of the translations themselves has also been blamed for reducing comparability (*ibid*), via both accuracy and repetition of terms in test items.

37. Factors of culture, language and translation are therefore important to comparability whether the respondent is answering test items, or responding to background questionnaires. To address these potential

forms of bias in child well-being data, Chapter 3 tests for cultural bias across selected indicators, lists items dropped from surveys altogether by countries due to national sensitivities (Norway, Poland, Turkey and the USA are examples of countries that dropped the questions on sexual activity at age 15 from their HBSC collections in 2005/06), and tests for bias from inconsistent translations in questionnaire texts over time.

1.4.3 Gaps in child coverage and the child well-being data

38. An important aspect of assessing the potential of surveys to inform national and international efforts to monitor child well-being is the identification of gaps in the indicators of child well-being and child coverage in the surveys. To provide robust information for policy-makers, it is important to be clear not only about which children the data refer to, but also what aspects of child well-being the data measure, and what is missing.

39. Chapter 4 reports the gaps in the coverage of child well-being indicators, by child age and specific indicators, provides vital information for future development of surveys designed to inform cross-national monitoring of children. Current cross-national surveys have little data on child protection, neglect or mental health issues. Moreover, the experiences of disabled children or data does not always cover all children of all ages (particularly in the early years), there are, for example, no cross-national data sources for health behaviours of children under age 10 (although some countries, for example Ireland, do deliver HBSC questions to children under age 10). Chapter 4, reporting the gaps in the child well-being indicator coverage, will aid the application of available data for monitoring the well-being of children, and provide recommendations for survey development.

40. Gaps will be identified by mapping child well-being indicators in the international surveys by broad dimensions based on dimensions of child well-being comparisons outlined in section 3.3.2 (socio-demographic, education and schooling, health and risks, income and deprivation, housing and environment, civic participation and time use, relationship indicators, and subjective well-being). Together these data will provide a reference point for the available cross-national indicators of child well-being by concept and child-coverage, and importantly where gaps in the knowledge lie.

41. Table 3 provides an example format of a matrix of indicators and gaps by child well-being dimension. Dimensions are broken down into sub-dimensions and mapped to key factors for explaining variations in outcomes such as: coverage by age, migrant status, and the manner in which the data is reported (scales or binary form, for instance).

Table 3: Mapping indicator availability and gaps: an example matrix for socio-demographics

	Age	Family form	Languages	Migrant status	Parental education	Religion	Sex
Number of indicators	7	19	6	26	15	2	6
Number of sources	6	9	3	7	4	2	4
Adult data (parents)	4	10	0	8	7	1	1
Child data	3	9	6	18	8	1	5
Child ages	9, 10 and 15	10, 11, 13, 15	9, 10, 13, 15	9, 10, 13, 15	13, 15	15, 16	9, 10, 13, 15
Migrant	7/7	17/19	6/6	26/26	15/15	2/2	6/6
Scales	5	11	2	3	1	0	0
Ordinal	1	0	0	4	9	0	0
Nominal	0	4	1	7	0	2	1
Binary	1	4	3	12	5	0	5

42. Chapter 4 will suggest ways to fill the gaps identified in this part of the project. And in doing so contribute to discussion in Chapter 5 that introduces recommendations for expanding surveys to children of difference ages (and identifying others sources of information), and experiences of surveys that sample students not in schools.

1.4.4 Recommendations for the use and improvement of international surveys of children

43. The main goal of this project is to provide methodological recommendations for the use and improvement of international datasets for informing the monitoring child well-being across countries and supplement national and transnational sources. Chapter 5 draws from the analysis undertaken in previous chapters to provide these recommendations and includes: basic recommendations are required for the calculation, reporting and application of the statistics derived, to make the best use of available data in international surveys of children.

44. In order to summarise the finding, and directly deliver recommendations, Chapter 5 is structured by the three project research questions:

- *What available data can be used to assess the well-being of children across countries?*
- *What are the strengths and weaknesses of existing international surveys for informing policy and monitoring the lives of children?*
- *How can the data for cross-national monitoring of child well-being be improved?*

45. Each of the preceding chapters will contribute independently to addressing the research questions in Chapter 5. The systematic review in Chapter 2 provides information on the methodology, content and coverage of surveys, setting boundaries for the evidence base they provide. The evaluation of the bias in the surveys in Chapter 3 will provide examples of methodological best practice for informing policy and monitoring the lives of children. A review of survey content in Chapter 4, from which a summary of gaps in the child well-being indicators organised by child characteristics and public policy dimensions is derived, will highlight data shortages, and missing dimensions and populations in cross-national knowledge base for monitoring child well-being.

CHAPTER 2. A SYSTEMATIC REVIEW OF INTERNATIONAL SURVEYS OF CHILDREN

Chapter 2 reviews the results of a systematic review of the surveys included in the report. It highlights information about the management structure, the data collected, the methods used, the logistics of collection, and the plans for future work. There are a number of common processes involved in the data collections, but often the methods applied vary. All surveys have expert consultants, defined and repeated topics for investigation and carefully account for representativeness of their collections and issues of comparability across countries. Moreover, all surveys have stated purposes which can cover issues such as well-being, and plans for analysis, such as inequalities and time trends analysis.

Using the results of this review, expectations for the extent to which these surveys can inform child well-being monitoring can be gleaned, and recommendations can be made in regards to good practise in the collection of child data for international comparisons and monitoring of child well-being.

2.1 Introduction

46. A starting point for assessing the availability and quality of comparative data of children is to review the surveys from which the data is derived. This chapter presents a systematic review of the six child surveys and three household surveys included in this report. [4] The aim of this review is to compare the surveys' management, content, and collections methods in order to inform decisions made for validity and robustness tests undertaken in the main empirical section of the report (Chapter 3). Results of the review will provide recommendations for good practise in, and set expectations for, the collection of comparable child data (Chapter 5).

47. After introducing the structure of the systematic review, the chapter addresses the parts of the review most relevant to the recommendations for use and improvement, presenting similarities and differences in the approaches of the different surveys.

48. Overall, results show that despite the common goal of surveys to maximise representativeness of samples (of pupils in schools, or households), different methods are being used, and different outcomes are being achieved. These differences may exist because of the content (achievement studies such as PIRLS, PISA and TIMSS require more detailed information), the result of the evolution of the surveys (academically driven surveys start small and use absolute threshold samples such as HBSC), or driven by different constraints, such as the finances available to the coordinators and participating countries.

[4] The review tables in Annexes 2 and 3 include more data than is covered in this chapter. The information in the Annex tables also reflects more recent information regarding content, coverage and methods based on updates in 2012.

49. What is evident however is that: although two of the surveys are designed specifically to capture well-being, each survey can uniquely contribute to our understanding of child well-being; funders could do more to complement their interests between the surveys; attempts could be made to align survey dates and countries covered; survey coordinators could learn from comparing the different approaches (for example, in terms of sampling methods, delivery in the field); and, there could be a greater engagement with the users of the data for the benefits of consistency across the waves on key child well-being indicators.

2.2 The structure of the systematic review

50. Information on each survey's management, content, and collections methods has been organised under the following themes: Management Structure, Data set basic information, Methodology, Delivery in the field, and Next wave/project. Under each theme, three to six subsections categorise the detailed information by theme to be compared and assessed. Table 1 shows the structure of the survey review.

Table 4: Structure of the systematic review of child surveys

Management Structure	Statement of purpose
	Funding and affiliations
	Co-ordinators and expert panels
	Government partners
Data set basic information	Survey dates
	Countries covered and target populations
	Contents (including core and contextual information)
Methodology	Unit of analysis
	Questionnaire development
	The sample design
	The sample threshold and treatment of under-sampling
	The timing of the data collection window
Delivery in the field	Training and management of field staff
	Translation and other procedures
	Length of tests
	Replacement sampling
	Additional modules are included
	Treatment of additional country specific items
Next wave/project	Next date and availability
	New content / questions if applicable
	Previous content not repeated in the next wave
	Countries added to, or leaving survey

51. To populate the sections of the review with information, the data, questionnaires, background information, and technical reports for each survey has been consulted. For the systematic review tables and survey coordinator details, please see Annexes 1, 2 and 3 of this document.

2.3 Management Structure

52. The first section reviews the management structures. The information collected here includes the statement of purpose, details of funding and affiliations, coordinators and expert panels involved with the studies, and government partners. In each case the management structure will contribute to the content and quality of the surveys, the extent to which it is used to inform research and policy decisions, as well as provide an idea of the flexibility of the study to expand or adapt to incorporate domains of child well-being not previously included.

2.3.1 Statement of purpose

53. Recent criticisms of comparative child well-being studies include the coverage (or otherwise) of key aspects of well-being, and the cross-sectional and aggregated form of analysis (for instance, well-being comparisons not accounting for variations within countries). The reason that much of this data/analysis is unavailable is because the cross national surveys, covering the countries of interest, do not collect such data (for instance data on child protection, neglect, or on children of different ages across a range of examples). The stated purpose of each survey included in this review shows what might be expected in terms of available data and coverage, both now and in future waves (and as such is vital to understanding present limitations in global data coverage).

54. The three education surveys (PISA, PIRLS and TIMSS) all purport to measure cognitive skills, yet differ slightly in terms how (life skills or curricula based, reading, mathematics or science), for who (by age or grade), and the contexts – and so the background data – by which the variation in cognitive skills might be explained. For instance the stated purpose of PIRLS is to monitor trends over time in reading literacy of children in middle childhood, with a particular focus on the factors, at home and in the school, which facilitate the acquisition of literacy in young children (PIRLS Assessment Framework and Specifications, 2006). TIMSS, although collecting educational achievement data from both the fourth and eighth grades (ages around 9 and 14, in middle and late childhood), is similar to PIRLS as it is designed to monitor trends in performance over time, and has a focus on home factors and school factors to address concerns about the quantity, quality, and content of instruction. An additional goal of TIMSS, with two age-cohorts, is to study the effectiveness of curriculum and instruction in relation to student achievement across times *and* across grades, as well as equity, for policy development through public accountability (See Annex 2 for source information).

55. The third educational survey, PISA, aims to measure the 'knowledge and skills essential for full participation in the knowledge society' for children coming up to end of compulsory school (age 15). The literacy measures include reading, mathematics and science tests that cover boarder life skills measures, rather than focus on curricula-based competencies. PISA, like TIMSS, explicitly states a design purpose to inform policy on the links between instruction and performance for children aged 15, by 'examining student performance within a single grade and gathering more data about classroom experiences'.

56. The first of the two health/risk behavioural surveys, ESPAD, is designed specifically to record alcohol, tobacco and drug use, in as many European countries as possible, to monitor time trends in alcohol and drug habits among students aged 16 in Europe, and to compare time trends between countries and between groups of countries. In the ESPAD statement of purpose, there is no explicit mention of informing policy. The second health/risk behavioural survey, HBSC, on the other hand, is designed to 'inform and influence health promotion and health education policy and practice for young people [aged 11, 13 and 15] at the national and international levels, and to advance scientific knowledge'. The study also seeks to understand the links between the life circumstances and social context and young people's health, health behaviour and lifestyles and build a better understanding of the factors that influence their well-being. And, as stated on the HBSC website (hbsc.org), 'data allow cross-national comparisons to be made and, with successive surveys, trend data is gathered and may be examined at both the national and cross-national level.' A number of items have been unchanged since the study started in 1983, and an explicit aim was to keep the mandatory part of the survey unchanged in three consecutive waves (2000/01, 2005/6 and 2009/10) in order to study time trends. In 2013, a journal supplement will be dedicated to reporting of time trends in health and wellbeing, behaviours and social contexts across countries that have participated in the three recent waves.

57. The stated purpose of ICCS is to investigate, in a range of countries, the ways in which young people are prepared and consequently ready and able to undertake their roles as citizens at age 14. To

understand how this varies at the national level, it will also collect and analyse data about student activities, dispositions, and attitudes related to civic and citizenship education, in the family in school context and at the national level. There is no explicit mention in ICCS documentation of informing policy, or measuring trends over time.

58. The three household studies were not designed with the explicit purpose of understanding children's knowledge, experiences or behaviours alone; and because of this their statements differ somewhat from the ones above. Moreover, the expectations for usable child well-being data derived from these surveys may be different from child-respondent surveys.

59. EQLS is designed for comparisons between countries and between demographic, social and economic groups, of household material conditions, employment situations, living and working conditions, family and community life, health and housing. In doing this, EQLS looks at the *views* of Europe's citizens on living conditions, their subjective well-being and their assessments of the society in which they live, and aimed at providing a useful contribution towards shaping the policies which seek to improve living and working conditions throughout Europe.

60. The stated purpose of ESS is to gather data about people's social values, attitudes, attributes and behaviour patterns within European polities. The academically driven household survey is intended to feed into key European policy debates through measuring and explaining the distributions of values, cultural norms and behaviour patterns, the ways in which they differ within and between countries, and how these results are trending.

61. EU-SILC is designed to be the EU reference source for comparative statistics on income distribution, living conditions and social exclusion at European level. The core purpose of EU-SILC is to contribute to the development of EU cooperation in social policy, and inform Member States in the development of their national (and sub-national) policies to tackle poverty and social exclusion. SILC allows the Member States and the European Commission to monitor national and EU progress towards key EU objectives in the area of social inclusion and of social protection, and to support mutual learning and identification of good (and bad) practices in terms of policies and institutional processes.

62. The most common stated purpose for each of these surveys is informing policy. Of all of the surveys, only ESPAD and ICCS do not clearly state informing policy as a key goal. However, well-being policies in particular, are a low priority: of all of the studies, only HBSC as a child survey, and EQLS as a household survey, explicitly mention well-being in their statements of purpose. In terms of possibilities for analysis, six of the surveys mention time trends analysis as a key purpose (PIRLS, TIMMS, ESPAD, ESS and EU-SILC, HBSC);HBSC has recently completed three identical waves in order to be able to analyse trends in as many countries as possible. It is also an important aspect of the national surveys and many countries which have been study members since the 1980s have more than two decades of trends (e.g. Norway, Scotland, and Wales). Also, it is important to note that two surveys (and only two) mention the 'comparison of distributions' as a key purpose (TIMSS and ESS), despite the fact that scaled micro-data available in many of the surveys allow for this type of analysis.

2.3.2 Funding bodies and affiliated organisations

63. Coordinating an international survey can be exercise in compromise with key stakeholders, or groups lobbying for change, and the core purpose stated for the survey at its inception. Demands from funders and affiliates are a central constraint on survey coordinators in developing the surveys, or trying to change (or retain) the structure or the content of the survey. Funders and affiliates represent the main stakeholders that are not the executors of the surveys, and can be transnational or national bodies, academic institutions, non-governmental organisations or government ministries.

64. Few of the surveys reviewed here have straightforward funding structures. The EU SILC survey and PISA are possibly the simplest examples, with the former being funded by the European Commission, and the latter receiving all of its funding via OECD countries for coordination purposes, and other participating countries for the remaining costs.

65. All other child survey coordinating centres collect participating costs from the countries included but the model for this varies. In HBSC country teams raise their own funds to conduct national surveys and they also pay an annual subscription that contributes to international coordination and data management. The payments for the PIRLS and TIMSS surveys are also annual (and so are paid in four or five instalments) and in recent waves have come to around 30 to 40 thousand USD per year per age or grade cohort (TIMSS countries with samples for both 4th and 8th grades pay for both, but at a discounted rate). ESS is the only household survey that receives funding directly from research bodies in participating countries.

66. A number of the surveys receive support from European organizations. ESPAD receives funding from the Pompidou Group at the Council of Europe, and has received support for working groups to test the ESPAD questionnaire in 2006 from the European Monitoring Centre for Drugs and Drug Addiction (a decentralised agency of the European Commission). The ESS is jointly funded by the European Commission and the European Science Foundation (and Independent body). EQLS is funded through the general budget of the European Commission.

67. Other regional organisations also contribute to the costs of the surveys. As mentioned above, the WHO Regional Office for Europe contributes to some research dissemination activities including international forums (http://www.hbsc.org/publications/forum/) and the publication of international HBSC reports. In the case of ICCS, CRELL (Centre for Research on Lifelong Learning an Institution of the European Commission) supports the regional module for European countries and the Inter-American Development Bank and CERLALC (*Centro Regional para el Formento del Libro en América Latina y el Caribe*) does the same for Latin American countries.

68. The IEA (International Association for the Evaluation of Educational Achievement), an independent organisation of researchers and government research bodies which oversees ICCS, TIMSS, and PIRLS (among other surveys) received support from various UNESCO sections for ICCS in 2009. The IEA itself supports PIRLS, TIMSS and ICCS via the IEA Data Processing Center in Hamburg, Germany, and the Educational Testing Service in Princeton, New Jersey, United States. TIMSS in 2007 and PIRLS in 2006 received support from the World Bank, the UNDP also contributed to TIMSS 2007.

68. In a number of cases, national government ministries and offices (research and statistical) can also directly contribute to the costs of delivering these projects at the international level. The Swedish Ministry of Health and Social Affairs has supported the ESPAD project, and some of the printing costs of the publication in previous years have been covered by the Swedish National Institute of Public Health (a subsection of the Swedish Ministry of Health). The main costs of international coordination HBSC have been met by NHS Health Scotland since 1995 when the current international coordinator (based in Scotland) was elected. National surveys are funded through a range of mechanisms including funding from government ministries, health agencies, scientific funding bodies, charities, and other sources. The HBSC International Coordinating Centre has been based in Scotland since 1995, and has received funding from the Chief Medical Office of the Scottish Government (previously Executive) and currently from NHS Health Scotland and member country subscriptions. For the 2006 and 2007 waves, a part of PIRLS and TIMSS funding came from the United States Department of Education through the National Center for Education Statistics.

69. It might reasonably be expected that when governments are part of the collection processes the results are more likely to be fed directly into policy discussions. Surveys with direct government involvement include ICCS (Ministries of Education of the participating countries); TIMSS and PIRLS (United States Department of Education); EQLS (the European Parliament, Norwegian and Swiss Ministries); ESS (European Commission); ESPAD (Swedish Ministry of Health); and EU SILC (the European Commission and the National Statistical Offices of EU member states).

71. Nongovernmental organizations such as universities can also be central to the funding of the surveys. For HBSC in 2009/10 (and previous waves), the Child and Adolescent Health Research Unit (CAHRU) (previously based at the University of Edinburgh and now at the University of St Andrews, Scotland) hosted the International Coordinating Centre; and the University of Bergen, Norway hosted the International Data Management Centre. Boston College, MA, USA contributes to the funding of PIRLS. And support in-kind, via expert panels on these surveys, are provided by many university across Europe and beyond (see Annexes 2 and 3).

72. What is striking is that on many occasions the funding sources can be the same for more than one survey. For instance, in the case of EU-SILC, EQLS and ESS, or in the cases of PISA, HBSC, or indeed any of the remaining surveys where government funding is used to meet the costs. In these cases it is in the interest of funders to make efficient purchases, and ask for harmonisation in the survey – even if the end goal is not for the purposes of child well-being – to link health and education surveys or living standard and social surveys, via harmonisation agreed at management levels through case (individual records where possible/ethical) or common item links (see Chapter 5 for further discussion). Such links would enrich the analysis available in both sets of data; and in turn provide better information for informing policy.

72. More often than not, the surveys are supported fully or in part by governmental organisations, and through public funds. Results of the surveys, therefore, should be applicable to policy and open for public use.

2.3.3 Co-ordinators and expert panels

73. The coordinators themselves, and the expert groups they employ or establish from the network membership in the case of HBSC, advise on the survey management and content and together are the key decision makers. These key decision makers can be important advocates of greater access and availability of child well-being data because they decide which items will be included, or dropped, from the questionnaires, which rotating modules will be undertaken, and which countries are accepted in future waves (as well as which countries might be encouraged to join). In cases where there are questionnaire development groups, as well as management consortium, recommendations, contacts and questions regarding developments to specific parts to the surveys can be delivered directly to these groups. The organisation and process of research development and decision making varies as described below.

74. ESPAD has a questionnaire group for developments and amendments. In recent waves the questionnaire groups have been informed by thematic groups. For example in 2004, four thematic groups were formed with around seven experts in each, covering the area of Attitudes, Gender, Risk groups and Trends. In 2010, there were seven thematic groups.

75. HBSC is coordinated by the elected International Coordinator who has established the study's International Coordinating Centre at the Child and Adolescent Health Research Unit (CAHRU), formerly at the University of Edinburgh until they moved to University of St Andrews in late 2011 with support of an elected Coordinating Committee and Assembly of Principal Investigators from each country. HBSC is a multi-disciplinary network of academics mainly based in Universities and Research Institutes throughout

the 43 member countries. Network members form expert groups (topic focus groups and scientific, methods and policy development groups), who work collaboratively to develop and produce each survey's International Survey Protocol including survey questionnaire, the content of which is decided together based on scientific rigour and policy relevance. Validation studies on questionnaire items are conducted during survey development.

76. HBSC international reports are also produced collaboratively, co-authored by Principal Researchers and Team members, with input from the members of the editorial group of the HBSC international research network. The authors work together in groups on topics based on disciplinary backgrounds and familiarity with related research paradigms. HBSC also employs an editorial board for international reports with elected leads from different groups, currently based in CAHRU, University of St Andrews, Scotland; the Karolisnkia Institute, Stockholm, Sweden; the Faculty of Social and Behavioural Sciences, Utretch University, Netherlands; the Welsh Government, and the Department of Health Promotion and Development, University of Bergen, Norway; and the WHO Regional Office for Europe, and the Social Services and Children analytical team in the Welsh Government. .

77. A joint management committee oversee the management of ICCS 2009. Included in these groups are three educational research groups from Australia, Italy, and the United Kingdom and the Secretariat of IEA. A project advisory committee consists of experts from within the consortium, CRELL, CERLALC and independent citizenship experts. ICCS also makes use of expert consultants at appropriate points in the study, selected on the basis of expertise, relevant to specific aspects of need.

78. PIRLS is coordinated by the IEA International Study Center (ISC) located at Boston College, United States. For the 2006 assessment, an update of the framework and specifications involved the Reading Development Group, the Questionnaire Development Group, and the National Research Coordinators. The National Research Coordinators are national representatives who are responsible for implementing PIRLS in accordance with international procedures.

79. PISA assessments are developed co-operatively, agreed by participating countries, and implemented by national organisations. The PISA governing Board (PgB), representing all nations at the senior policy levels, determines the policy priorities for PISA in the context of OECD objectives and oversees adherence to these priorities during the implementation of the programme. Participating countries take responsibility for the project at the policy level. Experts from participating countries also serve on expert groups for Science, Reading and Mathematics that set the PISA policy objectives with help from the best available substantive and technical expertise in the field.

80. As with previous TIMSS assessments, TIMSS 2007 (and 2011) has been coordinated by the IEA International Study Center (ISC) at Boston College, United States. Other members of the TIMSS consortium are the IEA Secretariat and the IEA Data Processing Center, Statistics Canada, and the Educational Testing Service, although the staff at TIMSS & PIRLS International Study Centre in Boston are responsible for the design and implementation of these studies. Expert groups involved in developing and producing the TIMSS frameworks with ISC include Science and Mathematics Item Review Committees and the National Research Coordinators of participating countries. Expert groups make recommendations on the content and cognitive domains, and focus areas for policy-orientated research. National Research Coordinators (NRCs) work with international project staff and review drafts of the frameworks (TIMSS 2007 Assessments Framework, 2007: 119-121).

81. The interview questionnaire for EQLS is developed by a research consortium. In 2003 the consortium consisted of academics from universities across Europe (including Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia, Slovenia, and Turkey).

82. ESS is led by the Centre for Comparative Social Surveys, City University, UK. Expert panels include the multi-national Scientific Advisory Board which comprises one representative selected by each national funding agency plus two representatives from the European Commission and the European Science Foundation respectively. This Methods Group advises on technical and methodological aspects of the survey, under the following panels: sampling panel specialists advises on specific issues regarding sampling and 'signs off' all national sample designs; and translation specialists responsible for guiding the translation process (guidance notes for and advice to National Co-ordinators).

83. EU SILC is coordinated by Eurostat, with the actual indicator development undertaken by an expert group, made up of national delegations of expert, under the responsibility of the Social Protection Committee (SPC) Indicators Sub-Group. The European Commission Directorate-General on "Employment, Social Affairs and Equal Opportunities" (overseen by EUROSTAT) makes up the secretariat.

2.4 Data set basic information

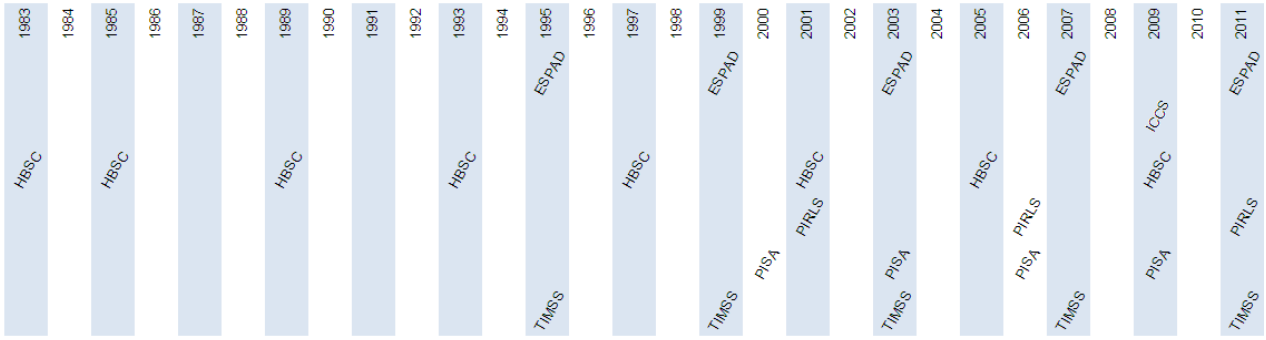
84. The second section of the review covers the basic information included in each survey's dataset. The information presented here includes the dates of the surveys (and the wave cycles), the countries covered and populations within these countries, and the contents of the survey (including both core and contextual items, and non-compulsory modules which countries may opt into as part of their national collections). The basic data provides details about what is being produced for comparing child well-being, and importantly when, for which groups of children and in which countries. In some cases this will include non-compulsory modules in the surveys which countries can opt into wave by wave.

2.4.1 Survey dates

85. Whether or not the information collected by each of these surveys can be used for reporting or monitoring child well-being will depend on the timing between updates of the data. Moreover, because international child well-being comparisons use multiple sources, how the waves coincide is of interest to researchers and policymakers planning to use the data, as well as for recommendations for harmonising collections.

86. The burden of organising, delivering, and processing an international child survey means that almost all of the child surveys are repeated on a four-year cycle (see Figure 1). The exceptions to this are PISA, which runs a three year cycle, and PIRLS which is undertaken every 5 years. The extent to which useful time series can be made, for developing time trends or pooled time-series macro-level analysis, can also be derived from this pattern – with the exception of ICCS each of the surveys have at least three completed waves. The most robust time trends by country (depending on repeated items and country coverage) can be derived from ESPAD, HBSC, and TIMSS. Finally, it is also useful to know the dates when the surveys were undertaken in the past in order for the data to be applied for use in natural policy experiments (either side of a policy change).

Figure 1: Dates of waves for the various international child surveys, 1983 to 2011



Source: Various child survey reports, websites. See Annex 2.

87. It is also worth noting that as the cycles are presently run, TIMSS and ESPAD, and HBSC will always be collected two years apart. Because PIRLS and PISA run in cycles of 3 and 5 years respectively, opportunities will arise to compare these surveys results with the results of all of the other surveys over time. This fact alone would suggest that efforts to harmonise surveys should concentrate on these two projects.

2.4.2 Countries covered

88. The countries covered in each survey will inevitably impact on the usability of the survey for cross-national comparisons in Europe and the OECD. The coverage of countries in the most recent wave was an important criterion for the inclusion of surveys in this project, and is presented in the executive summary. However, country coverage can also change between the waves, and depending on if the countries are repeated, time series can or cannot be provided.

89. Table 5 reports changes to country membership in the last three waves of each survey and projections for the upcoming wave. It is worth noting two important patterns. First, that all of the surveys are increasing in size over the years, meaning demands on schools in particular will be increasing, and efforts to innovate collection methods (when low to middle income countries are included – for an example see PrePIRLS) will be necessary over time. Second, a handful of countries are not permanent members of the collections, for example OECD countries of Greece, Israel and Turkey enter and leave the surveys. And some developed countries have never been part of unique health and risk surveys (ESPAD and HBSC), such as Australia, Korea, Japan and New Zealand.

Table 5. Survey country coverage and changes in the past three waves

	Countries covered[a]	Countries not repeating next wave	Countries to be added[c]
ESPAD – European School Survey Project on Alcohol and Other Drugs.	2003 and 2007, 35 Countries in total, including: Austria, Belgium, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, the Slovak Republic, Slovenia, Sweden, Switzerland, Turkey (2003 only), and the United Kingdom	Armenia, Austria, Switzerland Turkey did not take part in 2011	Additional countries in 2011 included: Albania and Liechtenstein.
HBSC – Health Behavior in School-aged Children	2000/01, 36 countries: Austria, Belgium (French, Flemish), Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Portugal, Russian Federation, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, the United Kingdom (England, Scotland, Wales), and the United States. HBSC 2005/06, 41 countries: The same countries adding Bulgaria, Iceland, Luxembourg, Romania and Turkey.	Bulgaria[b]	Additional countries in HBSC 2009/10 included: Albania and Armenia.
ICCS – International Civic and Citizenship Education Study	In 2009, 38 countries, including: Austria, Belgium (Flemish), Bulgaria, Chile, Cyprus, the Czech Republic, Denmark, Estonia, Finland, Greece, Ireland, Italy, Korea, Latvia, Lithuania, Luxembourg, Malta, Mexico, the Netherlands, New Zealand, Norway, Poland, Russian Federation, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland and the United Kingdom (England).	Not yet decided	Not yet decided
PIRLS – Progress in International Reading Literacy Study.	PIRLS 2001, 35 countries: Argentina, Belize, Bulgaria, Canada (Ontario, Quebec), Colombia, Cyprus, the Czech Republic, France, Germany, Greece, Hungary, Iceland, Israel, Iran, Italy, Kuwait, Latvia, Lithuania, Morocco, the Netherlands, New Zealand, Norway, Romania, Russian Federation, Singapore, the Slovak Republic, Slovenia, Sweden, Turkey, the United Kingdom (England and Scotland), and the United States. PIRLS 2006, 40 countries: including additionally Belgium (French and Flemish), Denmark, Luxembourg, Poland and Spain.	In 2006 Cyprus, the Czech Republic, Greece, Turkey all dropped out of the survey. None in 2011	Additional countries in PIRLS 2011 included: **Australia**, **Austria**, Azerbaijan, Botswana, Colombia, Croatia, (**Czech Republic**), **Finland**, Georgia, Honduras, Indonesia, **Ireland**, **Malta**, **United Kingdom** (**Northern Ireland**), Oman, **Poland**, **Portugal**, Qatar, Trinidad and Tobago and the United Arab Emirates.
PISA – Programme for International Student Assessment.	PISA 2003: 41 countries participated in the assessment: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Iran, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Russian Federation, the Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. PISA 2006: 57 countries, and to add to the above: Bulgaria, Chile, Estonia, Iran, Lithuania, Romania and Slovenia joined the survey.	None in 2009	Additional countries in PISA 2009 included: Albania, Argentina, Azerbaijan, Brazil, Colombia, Costa Rica (2010), Croatia, Georgia (2010), India, Indonesia, **Israel**, Jordan, Kazakhstan, Kyrgyz Republic, Liechtenstein, Peru, Qatar, **Romania**, Serbia, Thailand and Trinidad and Tobago.
TIMSS – Trends in Mathematics and Science Study	In 2003, 51 countries including: Australia, Belgium (Flemish), Bulgaria, Canada (Ontario and Quebec), Chile, Cyprus, Estonia, Hungary, Iran, Italy, Japan, Korea, Latvia, Lithuania, the Netherlands, New Zealand, Norway, Romania, Russian Federation, the Slovak Republic, Slovenia, Spain (Basque Country), Sweden, the United Kingdom (England, Scotland), and the United States (inc. Indiana State as a benchmarking participant). In 2007, 67 countries participated, including: As above plus Austria, Canada (Alberta, British Columbia, Ontario, Quebec), the Czech Republic, Denmark, Germany, Malta, Turkey and the United States (Massachusetts, Minnesota).	Indiana State (US) dropped out in 2007, as did Flemish Belgium, Chile, Estonia, Latvia dropped out in 2011	Additional countries in TIMSS 2011 included: Armenia, Azerbaijan, Bahrain, Botswana, **Chile**, Croatia, **Finland**, Georgia, Ghana, Honduras, Indonesia, **Ireland**, **Israel**, Jordan, Kazakhstan, Lebanon, Malaysia, Morocco, **United Kingdom (Northern Ireland**) Oman, **Poland**, **Portugal**, Qatar, Saudi Arabia, Serbia, South Africa, Syrian Arab Republic, Thailand, Tunisia, Ukraine, United Arab Emirates and Yemen.

EQLS – European Quality of Life Survey.	In 2003, 28 Countries, including: Austria, Belgium, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, El Salvador, Slovenia, Spain, Sweden, Turkey and the United Kingdom. In 2007, 31 Countries: as above including Norway.	None in 2011	The 2011 survey retained all of the 2007 countries with no countries dropping out.
ESS – European Social Survey.	2004/05, 25 Countries: Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and the United Kingdom. ESS 2006/07, 26 Countries: as above with Bulgaria, Cyprus, Latvia, Romania and the Russian Federation (see second column).	In ESS 2006/07 five countries did not participate that did in ESS 2004/05: the Czech Republic and Turkey.	In ESS 2006/07 four countries more are included compared to ESS 2004/05: Bulgaria, Cyprus, Latvia, Romania and the Russian Federation. Additional countries in ESS 2008/09 included: Croatia, **Israel** and **Lithuania**.
EU SILC – Statistics on Income and Living Conditions.	2005: EU member countries (with the exceptions of Romania and Bulgaria) and Iceland and Norway. 2006: as above. 2007: EU member countries (including Bulgaria and Romania), and Iceland and Norway. 2008: All EU member countries (with the exception of France), and Iceland and Norway.	None.	France returned to the survey in 2009.

Source: Respective survey datasets and websites.

Notes: [a]Countries listed in this table only include countries of focus in the study. Full lists of countries covered by the surveys are available in the annex of this document. [b]Bulgaria is still a member of HBSC, but was unable to complete the survey in 2009/10 due to problems with funding. [c]Countries in bold are European Union, or OECD, member countries.

2.4.3 Survey Contents

90. Understanding the content of the surveys is important for assessing the extent to which they can inform the comparison of child well-being, for understanding the types of analysis that can be undertaken with the data, and the extent to which the information can be use for informing policy beyond monitoring of indicators. From an indicators perspective, messages on the broad content will outline the extent to which the surveys can be used together to produce a monitoring tool or child well-being indices, and highlight additional information available from non-compulsory questionnaires that may be useful for exploratory analysis.

91. The information introduced in this section refers to the broad content of the surveys and not the content by item, which is covered in detail in chapter 4 and in the annexes of this report.

2.4.3.1 Core and contextual questionnaires

92. Core and contextual questionnaires are primarily differentiated in educational or civic studies that include a test questionnaire and background questionnaire. The core questionnaire is often delivered, in an exam setting using pencil and paper, and is a selection of items from a more comprehensive list of test items designed to provide enough observed response to estimate a 'literacy' level in that test. The purpose of the contextual questionnaires, or background questionnaires, in the surveys is to record the aspects of child's family background that might help to explain variation students' achievement.

93. ICCS for example, describes a two stage process with a core 'assessment' questionnaires covering a civic and citizenship framework with cognitive tests and a student's perceptions questionnaire; and background information on socio-demographics of students and their families, classroom environments, and teaching and learning providing the 'contextual framework'.

94. PIRLS takes a similar approach to ICCS, by including a core reading assessment using multiple choice and constructed responses (including reading for literary experience and reading to acquire and use information), and background questionnaires. PIRLS has four short background questionnaires including a student questionnaire of about 15-20 minutes (covering home and school experiences including basic demographic information, self perception, attitudes towards reading, computer use and literacy resources in the home), a home survey of about 10-15 minutes (covering child parent literacy interaction, parents reading habits and attitude, resources, home-school connections and socio-economic information), a teachers questionnaire of around 30 minutes (covering class characteristics, class resources and time, home-school connections, assessment practices and development opportunities) and a school questionnaire of around 30 minutes (covering including enrolment information, school and local area characteristics, socio-economic indicators of the student body, reading education characteristics and school resources and climate) (Mullis *et al*, 2009: 72-3).

95. The approach taken by TIMSS is similar to PIRLS, with the exception that the core questionnaire assesses student in mathematics and science (algebra, geometry, and data and chance in mathematics at the eighth grade and biology, chemistry, physics, and earth science in eighth-grade science, and since 2007 the cognitive domains of knowing, applying, and reasoning), and the background questionnaire is also administered to curriculum specialists, to parents from 2011, and from the children themselves more information is gleaned on curricula-relevant school experiences.

96. PISA, the final educational study, also undertakes the two-step collection method. The core questionnaire however rotates to focus on the domains of reading, mathematical and scientific literacy in turn. For the PISA achievement measure (main literacy test) about 390 minutes of test items are covered, with students taking different combinations of test items to a total of about 2 hours (in 2006). Pencil-and-paper tests are used, which include multiple-choice items and open-ended questions. PISA student background questionnaires take approximately 30 minutes to complete, and gather information on students and their family backgrounds (economic, social and cultural capital); students' attitudes towards learning, their school life and habits, and their family environment. Questionnaires are delivered to school principals also, and cover topics such as quality of the schools' human and material resources, public and private control and funding, decision-making processes, and staffing practices; context of instruction, including institutional structures and types, class size, and the level of parental involvement. In 2009, PISA also included a parental questionnaire for the first time that covers basic parent characteristics, child's past reading engagement, parents' own reading engagement, home reading resources and support, parents' background, parents' perception of and involvement in school, and school choice.

97. The two child surveys without a test questionnaire, ESPAD and HBSC, include socio-demographic information alongside the main measures of interest (health and risk factors). Included with socio-demographics in the ESPAD questionnaire are: alcohol, tobacco and drug related questions, leisure time use, educational attainment, and information about missing school. The HBSC core questions include: individual and social resources (body image, family support, peers, school environment); health behaviours (physical activity, eating and dieting, smoking, alcohol use, cannabis use, sexual behaviour, violence and bullying); and health outcomes (symptoms, life satisfaction, self-reported health, Body Mass Index, injuries).

98. Some of the child surveys include optional modules or questionnaires. PISA 2009 includes three optional cross curricular questionnaires on ICT skills, reading for school questionnaire, and an Educational Careers questionnaire. ESPAD has four modules covering the topics of Integration, Mainstream, Psycho-social measures, Cannabis, and Deviance. HBSC has optional modules that focus in more depth on the areas in the mandatory questionnaire. In a number of surveys, it is also possible for countries to include additional national-specific items in the questionnaire. This is the case for the main questionnaires of

HBSC and ESPAD, and for wealth items in TIMSS and PIRLS, and since 2003, PISA. In ESPAD, countries also have the opportunity to add their own modules (one or two).

99. Households surveys tend to consist of one core questionnaires and rotating modules that look at specific issues just once (or on a longer cycle – such as every four years). In the core questionnaire of each survey are a set of socio-demographic information such as: household size, employment status, level of education and marital status; however, the items used to collect this data can change in terms of wording and scaling.

100. Beyond the socio-demographic questions, the EQLS questionnaire examines a range of issues, such as employment, income, education, housing, family, health, work-life balance, life satisfaction and perceived quality of society. The face to face questionnaire consists of 183 questions items in total. EQLS has no modules.

101. The ESS questionnaire includes two main sections of similar sizes (approximately 120 items) a 'core' module and two or more 'rotating' modules, and a supplementary questionnaire is given to respondents at the end of the main interview, which in part, is used to evaluate the reliability and validity of items in the main questionnaire. In 2005/06 the two rotating modules were: Personal & Social Well-being: Creating indicators for a flourishing Europe, and The Timing of Life: The organisation of the life course in Europe. Each contained 50 additional items.

102. EU-SILC collections focus on income but also cover information on housing, labour, health, demography, education items as part of a multidimensional social exclusion framework. As with ESS, it contains a core annual collection as well as a rotating module containing 'target variables' (variables that, given EU SILC's principles of flexibility in implementation, may vary country-to-country) introduced every four years or less frequently. EU-SILC modules by year are: 2005: Inter-generational transmission of poverty; 2006: Social participation, 2007: Housing conditions, 2008: Over-indebtedness/Financial exclusion, 2009: Deprivation, and in 2010: Intra-household resource sharing.

2.5 Methodology

103. The third section of the review presents the methodology used in each of the surveys. In particularly this refers to the units of analysis, the data collection window, the sample design, thresholds and treatment of under-sampling. Information on the methodology in these surveys provides information on the coverage of the surveys, as well as a foundation for making assumptions about the representativeness of the surveys and the comparability between countries.

2.5.1 Unit of analysis

104. The group levels at which analysis can be undertaken on survey data is the unit of analysis. In complex surveys – like those undertaken with matching two-stage samples at the international level – this generally means that the pupil, school or the countries themselves can be units of analysis. In comparative household surveys this generally means the households and countries. The levels at which analysis can be undertaken defines that flexibility of the surveys for analysis of child well-being outcomes, and can also lead to recommendations for best practises in weighting and appropriate analysis. Each of the child surveys included in this report is undertaken in a school setting, although the sample methods for each survey can vary.

105. Both of the health and risk surveys report results primarily at the student level. ESPAD surveys students that turn 16 years old during the calendar year of the data collection. Students who are unable to understand or for other reasons cannot answer the questionnaire without assistance - e.g. children with special educational needs or severely handicapped - are not included in the ESPAD target population. The

HBSC international reports cover students aged 11, 13 and 15 years old. In neither study are school level results reported.

106. In ICCS, eighth grade students in participating countries are principal units of analysis. Units of analysis also include schools and educational systems. Specifically, target populations are young people in their 8th year of schooling provided that their average age is above 13.5 years of age (in the majority of participating countries this is the 8th grade).

107. For TIMSS, fourth and eighth grade students in participating countries are principal units of analysis, although the school or classroom can be units of analysis also. Target populations are the grades with the largest proportions of 9 year olds and 14 year olds in each participating country. The PIRLS survey samples students in Grade 4 (aged 9) as the principal units of analysis, as with TIMSS and PISA teachers and schools are sampled also. The principal units of analysis in PISA are students aged from 15 years and 3 (completed) months to 16 years and 2 (completed) months at the beginning of the assessment period – randomly sampled in the second stage of two sampling, the first of which samples schools (stratified proportionate to size of the school). In both PISA and PIRLS, primary care givers are also delivered questions as part of the collection.

108. EQLS surveys household heads aged 15 years and over, and resident in one of the 27 European Union countries, plus Norway, Switzerland and Turkey. ESS covers the same population, but is more precise and includes all persons 15 years and older who are resident in the country (regardless of nationality, citizenship or legal status). In EU-SILC two types of collection of household and individual variables are undertaken. In most countries, those without detailed registration data, all members aged 16 or more of selected households are asked to fill in a personal questionnaire. In countries with relevant registration information (DK, FI, IS, NL, NO, SE, SI), only a selected household respondent receives a personal questionnaire and household and income variables are collected either via register or the selected respondent.

2.5.2 The timing of the data collection window

109. A data collection window is used in international surveys in order to ensure that all of the international data is collected in a comparable time frame (generally within 2/3 months). The size of the data collection window may be used to infer comparability of results, with shorter windows leading to higher comparability.

110. When the window opens and closes, relative to the season of the year, is also an important consideration because school and work patterns in different countries are organised around seasons and national holidays. For example, the long school breaks in the year – generally preceded by end-of-year examinations – come in the summer, which for northern hemisphere countries is in June through to September and for southern hemisphere countries December through to March. These seasons can also affect working patterns in some households, or for some individuals, dependent on their job type. With the exception of the European surveys (ESAPD, ESS, EQLS, EU SILC) all other surveys include both northern and southern hemisphere countries.

111. In the case of most school surveys, differences in the school year in northern and southern hemisphere countries mean that collection windows are often split. This is to control for the additional learning time that children may have had in school years that start earlier when sampling by grade, and is the case for both TIMSS and PIRLS, for which data collection is conducted in Southern Hemisphere countries October-December the year before the survey year (so for PIRLS 2006, southern hemisphere countries were surveyed in 2005), and from March-June of the survey year for Northern Hemisphere countries. A similar collection window is used for ICCS 2009 (the survey took place October - December

2008 in the Southern Hemisphere and February - May 2009 in the Northern Hemisphere), and PISA where testing is conducted in a 42 day period between 1 March and 31 August (although PISA is sampled by age at the second stage).

112. In ESPAD, the only school survey confined to Europe, the recommended period for data collection is March/April in the year of survey. HBSC is not an educational survey and covers only northern hemisphere countries, nonetheless all fieldwork for each cross-national survey is carried out over a period of around seven to eight months, from October to May, to capture optimal sampling ages of whole school classes (11.5, 13.5, and 15.5 years). Each country has a limit of two months to carry out fieldwork in that period.

113. Household surveys, also produce a broad range of collection windows. For EQLS 2007 interviews were carried out in people's homes in September and October of that year, in 2011 this window was extended through to December. ESS also carries out data collection between September and December of survey year. And EU- SILC recommends that the survey fieldwork takes less than 4 consecutive months, with the time difference between the fieldwork and the income reference period for the sample being less than 8 months.

2.5.3 The sample design and thresholds

114. A final, and important, aspect of the methodology is the sampling methodologies. The design of samples to achieve the best representativeness of populations changes survey to survey, and can affect the treatment (adjustment weighting) required of final results to produce accurate estimates of population experiences across child well-being measures.

115. In recent years it has become clear that more and more surveys of children have put a great burden on schools (with national level surveys and in some cases school targets to achieve). In some instances this has led to a decline in first time response rates and a greater reliance on replacement samples to achieve representativeness in that country, and so overall comparability. In some recent international surveys, the failure of a replacement sample to achieve expected responses has led to a country being excluded from the final international comparisons in that survey. Therefore, sample sizes, and the level at which replacement rates are set, and whether they make a difference to representativeness, should be explored.

116. The goal of the sampling process in ESPAD to obtain a national and gender wise representative dataset, how this is done via sampling is up to each national researcher. There are central rules, however, such as the recommended number of participating students should not to be below 2.400, to allow for a breakdown of results by sex and another variable. And the final sampling unit in ESPAD is based on randomly-selected classes.

117. The sampling strategy of HBSC aims to achieve nationally representative groups of young people attending school at the mean ages of 11.5, 13.5 and 15.5. The sampling frame (the identified population from which the respondents will be sampled) should include 95% of the target population. Fifteen-hundred children per age group is the minimum recommended sample size, and so 4500 children per country. Samples may be stratified, and can vary country to country (school system, geo-political units, language – are the main stratifiers used). For the 2009/10 survey children were selected using a clustered sampling design, where the initial sampling unit was either the school class or the school selected based on probability proportionate to size. Schools were sampled when class lists were not available. To aid comparability, countries are given guidance notes for sampling, and sampling reports for each country are submitted and checked by – and feedback is given by – the Sampling Panel of the Methodology Development Group.

118. PISA sampling undertakes a two-stage probability proportionate to size sample. The first stage samples schools with 15-year old students by size of this population. Stratification methods are used, but can vary country to country. Per country, at least 150 schools are selected, but this can be much larger, with replacement schools being identified simultaneously form the school sampling frame (the national list of school including 15 year old populations). In the second stage, students are sampled in schools randomly until 35 students will be selected with equal probability (all 15-year-old students will be selected if fewer than 35 are enrolled). In 2003, a minimum of 4500 children per country were recommended, if children per school fell below 35, more schools were to be sampled. The number of students selected per school had to be at least 20, so between school analysis can be undertaken (an analytical objective of PISA) (School Sampling Preparation Manual PISA 2003 Main Study, Version One, June 2002: 7-8). In 2009, PISA tests were administered to between 5000 and 10,000 students, again from at least 150 schools in each country, and sample thresholds for inclusion in the final comparison were 85% of response school participation rates (before replacement) with student-level participation rates of 80%.

119. For TIMSS and PIRLS, as with PISA, probability proportionate to size sampling is used to select the schools. Unlike PISA, within school sampling randomly selects a class group (cluster sampling). For TIMMS, one of 14 student achievement booklets (each with part of the overall assessment) are assigned to pupils in such a way to ensure each booklet is answered by students of equivalent ability. Overall estimates can be produced because each question, or item, appears in two booklets, which links student responses between the various booklets. Item-Response Theory (IRT) scaling techniques are then used to produce overall estimates. A similar approach to reading literacy is taken by the PIRLS study. Both PIRLS and TIMSS set response rate thresholds at each stage of sampling. The example of PIRLS is 85 percent of both the schools and students, or a combined rate (the product of schools' and students' participation) of 75 percent. For TIMSS replacement schools are included, or the threshold under which comparability is deemed to be compromised, is 85%.

120. For ICCS, as with the previous educational surveys, a typical sample size for each country will be about 150 schools and between 3,500 to 5,000 students. Again, probability proportionate by size sampling is used for school selection; and like TIMSS and PIRL – the other IEA studies - within schools one classroom is randomly sampled to take the tests. The ICCS sample threshold is 85% of response school participation rates (before replacement), as well as student-level and teacher-level participation rates of 85%. Overall participation rates should be no lower than 75%.

121. For EQLS 2003 and 2007 a multi-stage stratified random sample was used which drew sample points from stratification by region and degree of urbanisation (random route method was used for clusters of 20 houses after the initial house was drawn at random). For larger countries sampling continued until around 1000 face-to-face interviews took place (in 2007 this was increased to 2000 for Germany and Turkey), and in smaller countries (including Cyprus, Estonia, Luxembourg, Malta and Slovenia) around 600 interviews took place (see Annex 3 for changes in 2011). For population estimates, weights are applied that inflated the sample according to age, gender and region within country.

122. The EU-SILC cross-sectional[5] data (whether survey or register based), are to be based on a nationally representative probability sample of the population residing in private households within the country, irrespective of language and nationality. The minimum effective sample size varies from a low of 3000 households (7000 individuals) in Malta, to a high of 8250 households (14500 individuals) in Germany. The plans are to achieve the minimum effective sample size of around 130 000 households or 2750 000 individuals aged 16 and over in the EU as a whole. However, given national over sampling in

[5] Information on longitudinal sampling is not included here. Please see Annex 2.

order to meet specific reporting needs the actual achieved global sample size is around 200.000 households.

123. Across the child surveys sampling methods vary. School-based sampling via stratification, and class and random sampling by grade and age are all used, as well as sampling for class lists directly when available. In terms of thresholds, absolute minimum headcount returns are required in all cases, (school count in PISA), as well as minimum response rates per school and student sampling before replacement (as in all of test-based studies). What is not completely clear in this discussion is the issue of excluded school and pupils form samples in all surveys. Exclusion can occur for various reasons, though most often active exclusion (students actively deselected from the sampling frame) are due to perceptions regarding student 'unable' to undertake the test for reasons of cognitive or functional disabilities. For example in ICCS exclusions are made in the case of three types of student: 1) Intellectually disabled students 2) functionally disabled students not able to perform in ICCS testing situation 3) non-native language speakers with less than one year instruction in the test language.

2.6 Delivery in the field

124. The fourth section of the review covers the methods by which the surveys are delivered in the field. Each of the surveys is a huge task involving many on-the-ground staff, the organisation of who will impact on the quality of the final product. Comparative surveys have the challenge of delivering the same survey in different languages in different school settings (depending on the age of the child as well as the system in which it is delivered) and so the training of field staff, and translations procedures for the questionnaires and support documents, are an important part of the quality assurance process. Sharing information on how these delivery challenges are met will provide important evidence on what makes for good practises, and in turn what can improve quality and comparability.

2.6.1 Central organisation, and organisation of field staff

125. How the training and management of field staff is undertaken by the survey coordination team is likely to impact on the quality of the surveys via efficiencies from improvement in field staff capabilities, and efficacies from expert guidance in a central coordination unit.

126. HBSC is run on a network model, involving experts from across all participating countries. The HBSC International Coordination Centre is responsible for coordinating all international activities within the HBSC research network, including: production of survey protocols and international reports, planning and organising the network's bi-annual meetings, facilitating communications, and acting as a resource centre for information on the study for members and external agencies and professionals. National teams train and manage field staff, when used, and an internal protocol provides guidance and documentation for use in schools to harmonise the instructions to field staff, schools and pupils as much as possible.

127. ICCS training and management of field staff is undertaken by The Australian Centre for Education Research (ACER), which serves as the International Study Centre (ISC) for the study. The ISC is also responsible for the overall design and implementation of the international test, student questionnaires, analysis and reporting. The 'parent' organisation of the ICCS, PIRLS and TIMSS surveys (and others), the IEA secretariat, manages the quality monitoring and translation verification procedures and communications with participating countries. National Research Coordinator (NRC) in each country, with the help of data entry and management staff and translation specialists are responsible for implementing the study in line with the international procedures.

128. Participating countries of PISA are responsible for the project at the policy level. PISA is implemented in each country by a National Project Manager (NPM) who is assisted by School Coordinators and Test Administrators.

129. The International Study Center (ISC) at Boston College serves as the International Study Center for IEA's studies in mathematics, science, and reading - TIMSS and PIRLS. Study Center staff are responsible for the design and implementation of the study, but support at the national level is provided by National Research Coordinators (NRC), who implement the studies in accordance with international procedures.

130. Household survey examples show a difference between efforts to manage the process of field organisations centrally, and the outsourcing of the management process. For instance, EQLS fieldwork in 2007 was carried out by a private company (TNS-Opinion, in 2011 Gfk EU3C carried out the fieldwork), which involved the employment of a network of national contractors, who were provided additional training specific to the survey, carried out the data collection in each country. In contrast, the ESS relied on the Central Coordination Team (CCT) to design and coordinate the survey – central coordination entails the participation from all responding countries, and independent advice from teams of independent academics.

2.6.2 Translation procedures

131. Good quality translation is an integral part of ensuring that respondents answer the same questions across countries, and that the responses are delivered without bias. Translation from a source language, and back-translation is commonly used to pick up potential inconsistency in word use across languages. Each international survey has translation procedures, although these vary in terms of the complexity and the management approach.

132. For different surveys translation methods can take the form of either recommendations or requirements. For instance, countries joining ESPAD, a translation-back translation of the questionnaire is highly recommended, as are field texts of the questionnaire. PISA on the other hand undertakes double translation from two source languages (English and French) for each language, with the original French version having been translated from the English 'via through double translation and reconciliation' (PISA technical report, 2009). In HBSC, all national questionnaires are translated from a written-English source, directly translated with adaptations allowed for 'linguistic clarity' (HBSC 2009/10 Protocol). Translations and back-translation are required, back translations are requested to be taken independently form translations, and an internal process flags up any items that have potential translation issues around meaning in particular countries.

133. In household surveys, examples of the management of the translation process include: decentralised and outsourced, and centralised. For instance, in 2007 translation for EQLS was carried out by a private research company employed to undertake the fieldwork (TNS Opinion) in conjunction with national contractors. ESS, on the other hand undertakes translation under the supervision of the Coordination Centre.

2.7 Next wave/project

134. The final section looks to the future and the plans for repeat waves in the variable surveys. When the next surveys will be undertaken and made available, if there will be new items included in the surveys, what items may have been dropped from the surveys; and finally whether countries have entered or left the surveys.

2.7.1 Next wave availability

135. One of the key criticisms of the attempts to monitor child well-being cross-nationally has been timeliness. In order to understand what is possible in terms of 'most recent data', and which surveys make their results available quickest, this section briefly look at the timing of next wave availability.

136. The dates for the next child surveys are outlined in section 2.4.1. The latest collections of HBSC and PISA were completed in 2009, for TIMSS, PIRLS and ESPAD this was 2011. For HBSC, 2009/10 survey data is the most recent available for analysis, but this requires researchers going through an application process. The HBSC study network and International Coordinating Centre are currently reviewing data access arrangements with a view to improving access to the micro-datasets. At the time of writing (March 2012), ESPAD 2011 was not yet available, as they were released in May 2012 with an extended report including more countries expected in early 2013. TIMSS and PIRLS, both 2011, are also not yet available.

137. Access to the data for the education surveys is more open, and despite containing complex test aspects, is generally available well in advance of the health and risks surveys. In each case, over recent waves, PIRLS, TIMSS and PISA data is made available at the end of the year following the collection year at the time of the release of the international reports (for example, in the case of PISA 2009 the full dataset was available for public access in late 2010).

138. The turnaround from collection to public release of household survey data is similar to that of the educational surveys. The most recently available EQLS data (2007) saw first results published in the second half of 2008. For EU SILC, both cross-sectional data and longitudinal sets are released 2 year after the collection, in March and August respectively (provisional versions). Finally, ESS 2006/07 was made available to registered researchers in September 2009.

139. Although the surveys release dates do not align, which is important for multiple-indicator monitoring tools or studies, it is clear that the delay on producing individual comparative child well-being indicators should be no different to producing income poverty statistics (which are rarely question for datedness). The difference seems to be due to some surveys (HBSC, ESPAD, and ESS), not only reporting later, but including application processes for access. These studies are also those coordinated outside of major transnational organisations (such as IEA, OECD, and EC).

2.7.2 Changes to content / questions if applicable

140. Due to the way in which questionnaires are developed, more often than not it is difficult to project changes in questionnaire content (or indeed specific questions). On some occasions, changes can be predicted due to rotation of the core topic of the surveys (as in PISA), or similarly know that changes are not going to take place as trends over time are being developed (HBSC).

141. The survey that has seen the most changes to background information over the waves is PISA, due to rotation of the core topic of the surveys. The main focus of PISA 2000 was on reading literacy, and it included an extensive set of tasks in this domain. In PISA 2003, the emphasis was on mathematical literacy and an additional domain on problem solving was introduced. PISA 2006 cycle focuses on scientific literacy. In each of these waves the background questionnaires have seen many changes to reflect the change in literacy interest, however changes have also been made to ever-present questions (including to the way in which questions on family structure have been asked and home wealth items).

142. In 2009 PISA returned to reading literacy, but again did not include identical background questionnaires to the 2000 study. For instance, information on the social interaction between parents and children (such as discussing schoolwork and eating meals together) was removed from the student

questionnaire, and placed in the parental questionnaire. To add to the parental questionnaire, PISA also include an Educational Career questionnaire for the first time in 2009, however these modules are not compulsory, and suffer low take-up rates at the national level.

143. Between the two most recently available waves, with the exception of PISA, few changes have been made to the survey content.

144. Two surveys have undertaken efforts to improve the data collection with limited or no changes to previous items. Between 2003 and 2007 a working group for ESPAD analyzed and reconstructed the ESPAD questionnaire in order to find methodological flaws that can produce questionable data, whilst maintaining the questionnaire as far as possible. The content of the questionnaire for each of the waves has remained mostly unchanged. In 2007, EQLS added new questions to the study, although all of the 2003 items remained unchanged (for 2011 changes please see Annex 3).

145. At the time of writing, the remaining surveys either did not change their core content, or were yet undecided. As expected, for time trends analysis, there have been very limited changes to the questionnaires for HBSC[6] (although there are plans to revisit the questionnaire for the 2013/14 wave) or PIRLS, and also for parts of the TIMSS survey (around half of the achievement items) for the same reasons. EU SILC has not changed its content either, although the survey does include rotating modules (for instance, in 2009 a module for child deprivation was included).

146. Despite a number of studies stating the purpose of developing time trends, it is often only when new data is published that researchers are aware of changes to the questionnaires. Although changes may be necessary for a range of methodological reasons, it would be useful for researchers and policymakers using these surveys to learn which items will be changed in advance of collection, and perhaps have a role in decision making on this front in order to help them plan future work and monitoring policies.

2.8 Chapter summary

147. This chapter has outlines important aspects of the management and methods of each survey, which will inform the following analysis and discussion. The results suggest that despite some specific differences in the ways surveys are run, common challenges and processes can be drawn between the surveys.

148. Among the more important findings are: the statements of purpose showing which surveys designed for well-being, those for trends analysis over time, and those for analysis of distributions; that funders can be involved with more than one survey; that governments across the globe are actively involved in supporting these works; and that the evolution of these important surveys in terms of content, country coverage and methodology continues. With this information in mind, it is hoped that the information outlined here can inform appropriate expectations of the surveys and use of their data, and perhaps lesson drawing between survey groups.

[6] Among the limited changes, some questions were added in risk behaviours to harmonise ESPAD questions, and two questions on neighbourhood factors were removed.

CHAPTER 3. AN EMPIRICAL EVALUATION OF INTERNATIONAL CHILD SURVEYS: ARE THEY RELIABLE?

Since the mid 2000s, there has been a flurry of multi-dimensional indicator-based comparisons of child well-being across the OECD. This chapter asks how reliable these indicators in these studies (and the underlying surveys) are for informing child and family policy decisions. The analysis reviews the data derived from surveys and, in particular, information reported by children themselves. Following a review of the survey data most commonly used in 12 recent child well-being comparisons, this paper introduces evidence fundamental to answering that question by exploring the survey data behind these indicators for evidence of inaccuracy, unreliability, and incomparability. Results raise fewer questions about the reliability and accuracy of the data per se, than the appropriate use of these data in cross-national comparisons. The chapter finds that, for all surveys, key areas for simple improvement in methods of indicator development can be identified. In conclusion, policymakers can be confident that appropriately-cautioned and appropriately-calculated indicators derived from children's surveys, are as reliable as equivalent indicators derived from adult surveys used to inform current economic and social policy.

3.1 Introduction

149. International surveys of children aim to target the same population of children across countries to answer the same set of questions, whilst ensuring comparability in responses through reducing system-based, methodological and cultural bias in the responses (Box 1). Bias in responses refers to differences between recorded values and true values imposed for reasons of exclusion of certain respondents from a sample, non-response, and incomparable responses due to cultural or social endogeneity in the questionnaire items.

150. The purpose of this chapter is to test for bias in child survey data used in the construction of child well-being indicators. Analysis of bias and associated methodological recommendations will be made from the perspective of child well-being indicators development only (for policy and/or advocacy). This means, for instance, that detailed considerations of data preparation for more complex analysis, such as regression analysis or forms of hierarchical linear modelling, will not be presented here. Moreover, analysis only covers data used to develop indicators for published comparative studies of child well-being, and will be applied to each of the nine surveys being analysed. Both descriptive and multivariate analysis will be undertaken – the latter will be performed at the national level and include controls for traditionally used sub-group measures of family type, gender, migrant status and rural / urban split (where available).

151. This chapter begins with an explanation of the possible effects of bias on comparisons of child well-being. Because the number of items in the average surveys questionnaire can run into the hundreds, it will then identify the main indicators used in recent comparisons of child well-being to select data for testing. The third part of the chapter introduces the methodology, before the fourth part presents the data quality tests and results organised by biases that are system based, sample or methodology based, and culturally or linguistically based (Box 1).

152. The summary conclusions of the evaluation will identify survey practises which have the greatest success in reducing or entirely avoiding biases of these types – and in doing so acknowledge good practise. Where biases are indentified, recommendations for improvements will be made. The results of this section will lead to recommendations for appropriate use of international data sources of children's well-being indicators for informing policy (Chapter 5).

3.2 Why worry about bias in international surveys?

153. The analysis of bias assesses the comparability, validity and reliability of indicators derived from the existing international surveys for informing policy and monitoring the lives of children. The problem with bias in country samples is that it affects the accuracy of the country estimates derived from the data, the reliability of trends for that country over time. With replicated bias, comparability and reliability can be ensured, if each country compared, suffers from the same degree of inaccuracy in comparable waves.

154. Undiagnosed bias in sample surveys leads to country estimates that can be inflated or deflated, giving a false impression of the position of the country in comparison with itself over time, or with other countries. The upshot of this inaccuracy is that policymakers with imprecise information could over- or under invest in a given policy, and the potential knock on effect of taking up time and resources that would otherwise be better used elsewhere. A longer term concern regarding the use of inaccurate data is the potential for generally undermining trust in child well-being indicators, or creating future problems – requiring further interventions and associated costs – as a result of misdirecting resources.

Box 1. Definitions of bias in the study

Analysis is this paper is organised in terms of bias tests. Taken together the bias tests will feed into broad assessments of practise and will feed into the recommendations to be outlined in the following chapter.

System-based bias refers to bias arising from the organisation of child and education policy in each country, and importantly the organisation of school systems, where many of the child surveys are undertaken. Education interventions for children including when children start school, the timing of the school year and examinations, children enrolled in mainstream schools (enrolment rates and drop outs) are all considered. Where country systems differ, there is potential for the survey process to introduce incomparability between the samples because of cumulative school experiences, or because more of the actual child population is captured in mainstream-school based samples.

Government policies affecting children's **actual** responses to surveys and not their **capacity to respond** (or to be included in the survey) are not forms of system-based bias. Government policies are treated as contributing factors to variation in true values (such as polices to provide school equipment that can affect deprivation scores).

Sampling and methodological bias refers to bias arising from the selection of sampling methods, questionnaire methodology, and response rates for schools and children that impose variation on national results. Population sizes vary much more widely than country sample sizes in all studies and in some cases the completion of a sample is decided once an absolute number threshold is passed. Questionnaires designed for the same purpose (such as measuring educational attainment, derivation, poverty or drug use) can be presented in very different ways in terms of length and structure. Tests for sampling and methodological biases will assess the extent to which these choices impact on indicator estimates, either via response rates to particular items, or accuracy in responses. Analysis of the weights designed for use with the survey will also be undertaken across indicators – by presenting weighted and unweighted results together where suitable – and adjustments to weighting systems will be recommended where appropriate.

Cultural and linguistic bias refers to bias that appears when national attributes and values affect the ability of children in that country to respond to the survey. Differences in translation methods and choices regarding wording, examples or items included in a survey are areas where such biases can occur.

In this analysis, the quality of data will be assessed in terms of:

- **no evidence of bias,** or evidence that
- **bias is systematic across countries** in respect of key items in the surveys used for comparing child well-being, or
- **bias is not systematic across countries** in respect of child well-being data and indicators.

Results with no bias will lead to recommendations for full inclusion of indicators in frameworks and countries in comparisons, with few or no cautions. Results where bias is systematic will lead to a recommendation for general reweighting of country results (or sets of results) and appropriate signalling of data issues in view of the reduced accuracy, but maintained reliability and comparability. Finally, results with non-system-based bias will lead to a recommendation for excluding relevant countries or waves of data from comparisons, or signal that validity, reliability and comparability are likely to be affected to varying degrees.

3.3 Analytical framework for assessing forms of bias

155. The analytical framework that will guide the empirical analysis of the data from each survey is outlined in Table 1. Tests will be presented in the order shown in Table 1, and include a description of the indicator. Only summary results will be presented in the main text, although complete results tables are available in the annexes of this document.

Table 6: Analytical framework for empirical exploration of various forms of bias

Bias	Indicators	Tests
System-based bias	Enrolment rates Dropout rates Starting school age	Cross-national descriptive and complex analysis; the latter including controls for gender and other factors. Tests to be included for between country differences in extent of bias. Analysis by selected child well-being indicators.
Methodological bias	Participation rates Questionnaire treatment (length, item order, item changes) Non-response rates	Cross-national descriptive and complex analysis; the latter including controls for gender and other factors. Tests to be included for between country differences in extent of bias. Analysis by selected child well-being indicators.
Cultural and linguistic bias	Translation in questionnaires Culturally sensitive questions (e.g. place of birth)	Review of translations in questionnaires by selected child well-being indicators, as well as changes to potentially culturally sensitive questions. Cross-national descriptive and complex analysis for bias including controls.

Source: OECD.

156. Each analytical section will record significances associated to measureable response differences, or significant differences in the socio-demographic attributes of non-response groups. Weighted and unweighted results will be compared side by side to show the effect of standard weighting procedures on the sample biases by indicator – and to recommend methods to adjust indicators for measurable response differences.

157. Some important similarities and differences in bias testing in previous literature and that to be undertaken here should be highlighted:

- Reviews of bias in surveys tend to concentrate on household surveys, or child surveys which require respondents to sit educational assessments (so education surveys). Examples of tests for bias in assessment surveys will be applied as appropriate to test for bias in non-assessment surveys (or background questionnaires) that ask children to provide self-reported data on behaviours and living conditions.

- Bias may or may not vary significantly between countries, which either way affects the accuracy of reported values used for simple cross-national reports and monitoring. However, more complicated statistical analysis (e.g. regression and trend analysis) may only be affected by bias if this significantly varies from country to country affecting the standardised (variance driven) results produced such analysis (Hanuschek and Woessmann, 2010, pp 9-10). Efforts to differentiate between these forms of bias will be made in all parts of the analysis, but the purpose of this work is only to inform indicator development and time trends.

- Much previous cultural and linguistic bias research has concentrated on identifying bias in one country, or by comparing results between two countries or regions. Cultural and linguistic bias is harder to identify across larger samples of countries which can translate questionnaires into multiple languages, and so it is not undertaken here, and new tests are required to cope with multi-country samples.

3.3.1 Selection of countries

158. The analysis will focus on the 34 OECD countries and the 6 non-OECD countries that are members of the European Union as of May 2011 (see section 1.3). These countries are: Australia, Austria, Belgium, Bulgaria, Canada, Chile, Cyprus[7,8], the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel[9], Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Malta, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Romania, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

3.3.2 Indicators selected for analysis

159. Each survey in this review has hundreds of possible usable items for child well-being indicator development, and so a selection of data / indicators for testing has to be made. Because this analysis looks specifically at indicators for child well-being, a review of the data used for child well-being indicators in recent comparative projects provides the foundation for selection.

160. Table 1 provides a comparison of frameworks and indicators from 12 child well-being studies (repeat studies share the same columns). The studies cover two global comparisons, four European-country frameworks, four OECD-country frameworks and a framework for central and Eastern Europe and the Commonwealth of independent states (CEE CIS – former Soviet countries). With the exception of the US Department of Commerce comparison from 1990, all studies have been undertaken in the past six years.

161. In each study, an attempt had been made to categorise data into clusters or dimensions based on policy interventions for children. Dimensions covering health, education, and economic/material status indicators were explicitly used in each. The number of individual indicators used in each study ranged from three, to over 70.

162. Table 7 categorises all indicators (regardless of how the authors originally categorised the data) into one of 10 dimensions: health and safety, material well-being and economic security, educational well-being, subjective well-being, personal family and peer relationships, behaviours and risks, housing and the local environment, quality of school life, social, economic and civic participation, and family forms and care.[10] The most commonly used indicators covering health (93 in total), behaviours and risk (61), and material well-being (47) – almost two-thirds of all indicators in total. After these three core clusters, educational measures are then most often reported (36), followed by housing and environmental indicators

---

7 Footnote by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognizes the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

8 Footnote by all the European Union Member States of the OECD and the European Commission: The Republic of Cyprus is recognized by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

9 The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

10 Each of these dimensions is covered in at least one of the studies. Differences from authors categorisations are made because minor differences in frameworks present crossover indicators that require re-categorisation.

(18), and social, economic and civic forms of participation (15). Indicators traditionally viewed as less objective (subjective well-being, self reported school experiences, and personal relationship measures), were used the least – ranging from ten to 14 times in total.

163. Not all of the data used in the studies were derived from surveys of children or households with children. Of the 326 indicators listed, 118 are from administrative sources, and commonly include mortality data, health services for young children (low birth-weights, immunisations, etc.) and educational enrolment amongst other aggregate statistics. The remaining 208 indicators were derived from surveys including: the European School Project on Alcohol and other Drugs (ESPAD), the European Survey of Income and Living Conditions (EU SILC) the European Quality of Life Survey (EQLS) and the European Social Survey (ESS), the Health Behaviour in School-aged Children (HBSC) study, the Progress in International Reading Literacy Study (PIRLS), the Programme for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS).

164. Table 8 outlines the most commonly used indicators in the projects that are derived from survey data only. In the vast majority (with the exception of the entirety of UNICEF Report Card 9 that looks at inequalities) these indicators are averages, rates or other forms of efficiency aggregates, and represent the average child without breakdowns by sex, age migrant stats, or social gradient (OECD, 2009 does do some of this however).

165. The top 14 indicators (all those that have seven or more repeated uses across all frameworks) clearly identify education, income poverty and material deprivation, and health and risk behaviours as important indicators to these frameworks. Relationships with peers and families are covered widely when used (included in four studies but using 12 indicators), as are subjective perceptions of children's well-being across a range of personal contexts.

166. Selections from the 14 indicators will be tested for forms of bias in the analysis below. To add to the standard tests – where possible – a range of socio-demographic indicators will also be factored into the analysis, including: age, sex, migrant status, rural / urban community, and parental education. At least one indicator from each of the surveys, and each of the dimensions of child well-being, will be covered in the analysis.

Table 7: An overview of indicators used in comparative child well-being analysis

	European multidimensional indices (Bradshaw et al 2007 and 2009)[1]	Unicef Report Cards 7 (2007) and 9 (2010)[1]	CEE CIS index (Richardson et al, 2008)	Doing Better for Children (OECD, 2009)	Save the children (2008 and 2011)[1]	European Commission Child poverty and Well-being (2008)[2]	European Commission TARKI (2011)	US Dept. of Commerce (Hobbs and Lippman, 1990)[3]	OECD Family Database (since 2007)
Health and safety	Child health from birth (infant mortality; low birth weight). Health behaviour (dental; nutrition x2; physical activity; BMI). Mortality rates (all, 0-19, suicide). Vaccinations (measles; DPT3; Pol3)	Child health from birth (infant mortality; low birth weight). Health behaviour (nutrition x 2 and [dist., 2010]; physical activity [dist., 2010]; BMI); Health Problems (morbidity [dist., 2010]); Mortality rates (all, 0-19). Vaccinations (measles; DPT3; Pol3)	Child health from birth (infant mortality; low birth weight; breastfeeding x 2). Mortality rates (under 5; accidental). Vaccinations (DPT3; Pol3; measles). Health Problems (morbidity, respiratory; morbidity, diarrhoea; dental; stunting; underweight; wasting). Health behaviour (nutrition; physical activity)	Child health from birth (low birth weight; infant mortality; breastfeeding). Health behaviour (physical activity). Mortality rates (all, 0-19; suicide). Vaccinations (pertussis; measles)	Health Problems (BMI); Mortality rates (under 5)	Child health from birth (breastfeeding[c]; life expectancy[a]; infant mortality[b]; low birth weight[b]). Health behaviour (nutrition[b] - breakfast; nutrition[c] - 5 a day; dental[c] x 2; physical activity x 2[b,c]). Health Problems (BMI[b] and [c]; mental health[c]; morbidity - chronic[c]; morbidity - infectious[c]). Mortality rates (suicide[b]; suicide and self-harm 10-24[c]). Vaccinations[c]	Child health from birth (breastfeeding; low birth weight; infant mortality). Health behaviour (dental; nutrition x 2). Vaccinations	Child health from birth (AIDS transmission; infant mortality x 2 [inc. first day rate]; low birth weight). Health Problems (morbidity); Mortality rates (suicide; male youth motor accidents)	Child health from birth (breastfeeding; infant mortality; life expectancy; low birth weight). Health Problems (morbidity - diabetes and asthma; BMI). Mortality rates (suicide). Vaccinations
Subjective well-being	Subjective well-being (life; peers x 3 [2007]; health).	Subjective well-being (health; life; peers x 3)				Subjective well-being (home[c]; school safety[c] x 2)	Subjective well-being (life)		
Personal family and peer relationships	Family relations (mother, 2009; father, 2009). Peer relations (classmates)	Family relations (meals; talking). Peer relations	Family relations (x 2). Peer relations (x 2).			Peer relations x 2[b,c]			
Material well-being and economic security	Income poverty (relative; gap). Deprivation (household x 2 [one in 2009]; educational items). Economic strain (2009). Jobless households	Average disposable income [dist., 2010]. Income poverty (relative; household x 2; educational items and [dist., 2010]). Jobless households	Income poverty (absolute). Deprivation (educational items x 2; household)	Average disposable income. Income poverty (relative). Deprivation (educational items)		Income poverty (relative[a]; relative gap[a]; persistent[a]; in-work[a]; anchored[a]; minimum income threshold[c]; absolute[c]). Jobless households[a] (x 2). Economic strain[b]. Deprivation (educational items[b]; environment[b]; household[b]; ICT outside school[c]; services[c] x 2; services[b]; care services[b])	Income poverty (relative x 2). Deprivation (educational items; household x 2 [inc. severe forms])	Average disposable income; Income poverty (absolute x 2; poor earnings; poor transfers)	Average disposable income. Income poverty (relative)
Educational Well-being	Educational achievement (reading; mathematics; science). Educational participation (youth; pre-primary). Aspirations (2007)	Aspirations. Educational achievement (reading; mathematics; science and [dist., 2010]). Educational participation (youth)	Educational achievement (reading; mathematics; science). Educational participation (pre-primary; primary; secondary).	Educational achievement (mean literacy; distribution)	Educational participation (primary)	Educational achievement (dist. in reading skills[c]; literacy[a]). Educational attainment (early leavers[a]; failure rate in lower secondary[c]; pass rates secondary[c]; failure rate in secondary[c]). Educational participation (truancy[b])	Educational achievement (literacy at 10; literacy at 15)	Educational participation. Educational attainment (x 2)	Educational attainment (years in; graduation rates). Educational achievement (literacy at 10; literacy at age 15)

Table 7: An overview of indicators used in comparative child well-being analysis (cont.)

	European multidimensional indices (Bradshaw et al 2007 and 2009)[1]	Unicef Report Cards 7 (2007) and 9 (2010)[1]	CEE CIS index (Richardson et al, 2008)	Doing Better for Children (OECD, 2009)	Save the children (2008 and 2011)[1]	European Commission Child poverty and Well-being (2008)[2]	TARKI (2011)	US Dept. of Commerce (Hobbs and Lippman, 1990)[3]	OECD Family Database (since 2007)
Behaviour and risks	Risk behaviour (smoking; alcohol; drugs, cannabis; drugs, inhalants, [2007]). Sexual health (Fertility 15-19; active <15; safe sex). Violence and violent behaviour (fights; bullying)	Risk behaviour (smoking; alcohol; drugs). Sexual health (pregnancy 15-19; active <15; safe sex). Violence and violent behaviour (fights; bullied)	Crime (juvenile rate; petty). Risk behaviour (smoking; alcohol; drugs x 2). Sexual health (Fertility 15-19; safe sex x3; active <15). Violence and violent behaviour (bullied x 2).	Risk behaviour (smoking; alcohol). Sexual health (Fertility 15-19). Violence and violent behaviour (bullied)		Crime (offered drugs[c]; offending[c]; victimisation[c]). Risk behaviour (accidents x 2[b,c]; alcohol[c]; smoking[c] x 3; drugs[c]). Sexual health (pregnancy 15-19[b]; fertility 15-19[c]). Violence and violent behaviour (bullied[c]; experiences of[c])	Risk behaviour (smoking; alcohol x 2; drugs - illicit; drugs - medicinal)	Sexual health (abortions 15-24; pregnancy 15-19; fertility 15-19); Violence and violent behaviour (mortality)	Risk behaviour (smoking; drugs)
Housing and the local environment	Environment (crime; dirt etc., 2009). Housing problems. Overcrowding	Overcrowding [dist., 2010].	Overcrowding. Housing problems (sanitation; water)	Overcrowding. Environment (dirt etc.)		Environment (climate[c]; dirt etc.[c]; play space < 5[b]; space for activities[c]). Housing problems[b] (comfort[b]; basic standards[c]). Overcrowding (space[b]; % of children[c]; own room[c] x 2)			
Quality of school life	Well-being at school (pressure; enjoyment)	Well-being at school		Well-being at school (enjoyment)		Well-being at school[c]. Deprivation (education services[b] x 2; ICT at school[b]; school health services[c]; school food services[c])			
Family forms and care		Family form (single families; step families)	Family form (divorce). Child discipline (x 2). Children in care (foster; residential; infant homes).			Family form (break-ups[c])		Family form (early marriage; single youth; median age - marriage; median age - first birth)	
Social, economic and civic participation	Civic participation (x 2 in 2007). Youth employment (NEET)	Youth employment	Child labour	Youth employment (NEET)		Civic participation (social[b]). Youth employment (NEET[c])		Youth activity (employment x 2 [inc. by sector]; unemployment x 3)	Civic participation (voluntary work; voting). Youth employment (NEET)

Table notes: (1) Indicators included in both studies have not been dated. “dist” refers to indicators presented at distributional measures. Indicators marked ‘x 2’ means multiple indicators of that type are included in the framework. (2) Indicators from the European Commission report are recorded with a postfix of 'a', 'b' or 'c'. Indicators with an 'a' are those based on commonly agreed EU indicators. ‘b’ are not common EU indicators, but could be available across EU sources, and only used by some EU countries. ‘c’ are alternatives to EU indicators used in some countries for which EU sources are not suited (EC, 2008: 78). (3) Indicators exclusively falling outside the age range of children (0-17) have been excluded, as have demographics indicators included in the publication.

Table 8: The most commonly used survey-based child well-being indicators

Indicator	Uses
Educational achievement (literacy at 10 and 15, mathematics, reading, science)	20
Income poverty (absolute, gaps, relative)	18
Deprivation (educational items, household)	16
Subjective well-being (health, home, life, peers and school safety)	14
Peer or family relationships	12
Health behaviour (nutrition)	9
Risk behaviour (hard drugs, Cannabis, inhalants, medicinal)	9
Risk behaviour (smoking)	9
Violence and violent behaviour (bullied, bullying and fights)	9
Health Problems (BMI, mental health, morbidity)	8
Overcrowding	8
Sexual health (active <15, safe sex)	8
Health behaviour (physical activity)	7
Risk behaviour (alcohol)	7

Table note: commonly used indicators are those used 7 or more times across the child well-being frameworks.

3.4 Analysis of bias and examples from the literature

167. This empirical section reviews examples of previous efforts to explore forms of bias in the surveys, before going on to perform tests on the data itself. The discussion is around three types of bias: system-based bias, sample and methodological bias, and cultural and linguistic bias.

3.4.1 Analysis of system-based bias

168. To explore whether system-based biases exist in the international surveys, the children enrolled in mainstream schools, the rate of drop-outs in the salient child cohorts, and the age at which children start school will analysed for their affect on key survey results.

169. An example of system-based bias testing comes from Hanuschek and Woessmann (2010) who analysed PISA data for an effect of the rate of school enrolment in the child cohort tested by PISA on average national test scores. Because all of the child surveys reviewed in this study are delivered in school or classroom settings, results reflect school populations and not strictly child populations. One important systematic form of bias in results therefore is introduced by enrolment rates of children in mainstream schooling at the time of the test (so excludes truants, long-term absent or those in schools excluded from the survey samples [such as special needs schools]). The authors conclude that countries with higher overall school enrolment for the cohort being tested perform better on the PISA league tables. A finding that runs contrary to the assumption that low enrolment, and therefore selectivity in sampling favouring those children with school ambitions or abilities, would be reported alongside higher average test scores. When compared with rates of exclusion from the surveys and non-response rates, significant associations for enrolment rates are found more often, showing positive correlations with tests scores in PISA 2000, 2003 and TIMSS 2003 (*ibid*, p6).

170. Analysis undertaken by Hanuschek and Woessmann (2010) suggest that researchers attempting to test for system-based bias need to access data on enrolment rates in schools for the test cohort, and preferably link these figures to exclusions based on survey selection criteria as opposed to all children 'out-of-school' as a group.

3.4.1.1 Indicators and relationships with dropout rates and enrolment rates

171. In order to assess the extent to which enrolment rate in schools and dropout rates (by sex) between cohorts might affect the final results reported in any given survey, surveys data over waves is pooled in time series with information on the target population (enrolment and dropout rates) derived from education accounts. Below Tables 9 and 10 report the correlation analysis and sensitized results and show, in PISA, rates of enrolment are closely associated to average reading literacy scores as well as deprivation scores over the waves.

172. Three PISA variables are included in the analysis. The first is reading literacy achievement. PISA collects literacy scores for reading, science and mathematics. Because the first wave of PISA in 2000 covered reading as the focal literacy measure, and this was repeated in 2009 (in 2003 mathematics was the focus, and science in 2006), reading literacy tests have been used most often, and are chosen for analysis here. Reading literacy results are explored in terms of mean values (most commonly reported), the spread of results within each country using a ratio of the 90th percentile result over the score at tenth percentile (used in both OECD, 2009 and UNICEF 2010) and the range of the mean standard error in 2000, 2003, 2006 and 2009.

173. The second PISA child well-being indicator is educational deprivation scores (used in no fewer than 5 of the 12 child well-being studies identified above). Because the number of items available for comparison over the PISA waves have changed (e.g. the background questionnaire no longer includes a question about the ownership of a calculator) a new seven-point scale of educational deprivation has been calculated for each of the four waves (a desk to study, a quiet place to work, a computer for school work, educational software, an internet connection, a dictionary, and school textbooks). Aggregated results are presented after applying the final student population weight in each wave.

174. The third PISA variable records children's aspirations for a job at 30. This is an open-ended question which asks children what job they expect to have at 30, and is categorised using Ganzeboom's (1992) international socio-economic scale. The item was used to classify children by skills aspiration in UNICEF's first report (2007) and in Bradshaw *et al*'s European Index in 2007. The measure has since been discontinued by PISA, after it was excluded from the compulsory questionnaire in 2003[11] and reinstated in 2006.

175. System-based bias testing of these outcome indicators uses enrolment rates and dropout rates (by gender). Enrolment rates for the 15-year-old populations have been calculated using statistics derived from the OECD population and education databases (downloaded in July of 2011). The dropout rates are calculated using the ratio of gross enrolments by sex, for the same cohorts, the year before and the year of the PISA tests (so gross enrolment of children aged in 2009 over gross enrolment of the same cohort in 2008 when aged 14). Although imperfect, it represents a rate estimate for children who dropped out of the school in the year of the PISA net of migration and immigration.

176. Table 9 presents the results of the correlations for the complete sample, as well as pooled results after sensitivity tests (removal of countries with enrolment rates below 90%). At first glance, higher levels of enrolment are strongly associated with both higher PISA mean values (and so in line with the Hanuschek and Woessmann, 2010) and lower levels of educational deprivation and future work aspirations. Weaker associations are seen for distributional measures of 90/10th percentiles in reading literacy, and the standard deviation for job aspirations, although directions of association follow those of

[11] Countries that did not complete the question in 2003 included: Canada, Denmark, Finland, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland and Turkey.

similar indicators above. Together these results, contrary to common conception that higher enrolment may mean lower dropouts of difficult or disadvantaged students, suggest that higher enrolment is strongly associated to higher well-being measures across the PISA waves.

Table 9: After sensitivity testing, enrolment rate associations weaken and most significant dropout rate associations disappear

Pooled time series associations (2000, 2003, 2006 and 2009) before and after sensitivity tests for PISA well-being indicators, enrolment rates and dropout rates by gender[12]

	All countries				Countries with enrolment rates over 90%			
	Enrolment rate	Drop out rate at 15 (boys)	Drop out rate at 15 (girls)	Drop out rate at 15 (all)	Enrolment rate	Drop out rate at 15 (boys)	Drop out rate at 15 (girls)	Drop out rate at 15 (all)
Mean PISA reading literacy	.623**	.518**	.549**	.534**	.285**	0.161	0.169	0.172
90/10 percentile ratio	-.202*	-.220*	-.210*	-.200*	-.316**	-0.149	-0.080	-0.123
Reading SE range	-.196*	-0.143	-0.129	-0.139	-0.190	-.204*	-.205*	-.228*
Educational deprivation	-.731**	-.512**	-.519**	-.525**	-0.159	-0.050	0.060	-0.008
Aspired employment at 30 (ISEI scale)	-.583**	-.486**	-.475**	-.497**	-.289*	-0.128	-0.150	-0.147
Aspired employment at 30 (ISEI scale - STDEV)	.508**	.337**	.266*	.328**	.399**	0.211	0.147	0.192

Note: Analysis of two-tailed Pearson correlations using pooled PISA data for four waves in 2000, 2003, 2006 and 2009. Enrolment and Dropout rate figures are calculated using data taken from the OECD Education database (2011). ** and darkest shading denotes significances at the $p < 0.01$ level, * and lighter shading denotes significances at the $p < 0.05$ level. Dropout rates are reversed.

Source: OECD calculations, OECD PISA (various waves), and OECD Education at a Glance (2011).

177. Because the range of gross enrolments of 15 year olds within Europe and the OECD vary massively (between 102.2[13] in Ireland and 63.7 in Turkey in 2009) – and the widening coverage of PISA now includes countries with traditionally earlier school-leaving ages – sensitivity testing of these correlates was undertaken to see if outliers in enrolment rates affected the results. Countries with lower than 90% enrolment rates were removed from the pooled tests including: Mexico and Turkey in all waves, Portugal in 2003 and 2006 and Luxembourg in 2006 (this left an N of above 100 for each correlation with the exception of tests for aspirations because there is no data for 2009).

178. The right hand side of Table 4 shows that many of the previously significant results have changed, either weakening in the case of most enrolment rate estimates[14] or disappearing entirely from dropout rates estimates for all but one of the outcome measures. There are now no significance reports for deprivation scores. After sensitivity test associations only three well-being indicator correlations strengthened. The size of the error range in reading literacy is now marginally significantly associated to dropout rates, suggesting that in high enrolment countries with fewer dropouts also have a lower variation in reading literacy results. A result that is supported by an increase in the strength of the association between the percentile range ratio and enrolment rates after sensitivity testing.

12 Data is available on numbers of special needs children for most European countries. This information will be incorporated into the tests in the coming weeks.

13 Gross enrolment figures can be above 100% when enrolment exceeds population due to net migration.

14 Scatter plots to be added to the annex.

179. Table 10 breaks down the same results by wave, and shows a similar pattern of significances that disappear once low enrolment countries are excluded. Pre-adjustment biases are most significant for reading in 2003, for deprivation in 2009, and for aspirations in 2000. After removing Mexico and Turkey from the 2000 wave's tests, high enrolment and low dropout rates overall (and for boys) remain significantly associated with higher reading literacy means scores. In 2009 with the same countries excluded, enrolment rates correlate strongly with higher mean reading scores and lower distributions in reading scores.

Table 10: Strong associations between enrolment and dropout rates do not hold for high enrolment countries

Associations, before and after sensitivity tests, for PISA well-being indicators, enrolment rates and dropout rates by gender, 2000, 2003, 2006 and 2009.

		All countries				Countries with enrolement rates over 90%			
		Enrolment rate	Dropout rate at 15 (boys)	Dropout rate at 15 (girls)	Dropout rate at 15 (all)	Enrolment rate	Dropout rate at 15 (boys)	Dropout rate at 15 (girls)	Dropout rate at 15 (all)
2000	Mean PISA reading literacy	.550**	.548**	.532**	.532**	0.294	.525*	0.370	.511*
	90/10 decile ratio (literacy)	0.000	-0.104	-0.174	-0.055	-0.385	-0.260	-0.182	-0.231
	Educational deprivation	-.757**	-.495**	-.453*	-.480**	0.050	-0.251	0.143	-0.057
	Standard error range (literacy)	0.004	-0.127	-0.121	-0.117	-0.225	-0.328	-0.328	-.420*
	Aspired employment at 30 (ISEI scale)	-.549**	-.500**	-.521**	-.554**	-0.130	-0.171	-0.354	-0.301
	Aspired employment at 30 (ISEI scale - STDEV)	.648**	.531**	.365*	.518**	0.147	0.282	0.138	0.235
2003	Mean PISA reading literacy	.744**	.491**	.587**	.541**	0.112	0.055	0.070	0.065
	90/10 decile ratio (literacy)	-.422*	-.421*	-.440*	-.439*	-0.199	-0.222	-0.196	-0.223
	Educational deprivation	-.795**	-.472**	-.558**	-.520**	-0.162	-0.055	0.016	-0.038
	Standard error range (literacy)	-.644**	-0.283	-0.374	-0.332	-0.356	-0.106	-0.143	-0.137
	Aspired employment at 30 (ISEI scale)	-.551*	-.523*	-0.419	-.475*	-0.242	-0.151	0.046	-0.055
	Aspired employment at 30 (ISEI scale - STDEV)	0.109	0.133	0.124	0.129	0.438	0.227	0.245	0.239
2006	Mean PISA reading literacy	.568**	.551**	.529**	.543**	0.255	0.209	0.136	0.180
	90/10 decile ratio (literacy)	-0.200	-0.238	-0.171	-0.207	-0.256	-0.128	0.041	-0.051
	Educational deprivation	-.732**	-.591**	-.590**	-.594**	-0.361	-0.169	-0.133	-0.157
	Standard error range (literacy)	-0.165	-0.131	-0.102	-0.118	-.379*	-0.305	-0.205	-0.265
	Aspired employment at 30 (ISEI scale)	-.636**	-.522**	-.514**	-.521**	-.446*	-0.125	-0.055	-0.095
	Aspired employment at 30 (ISEI scale - STDEV)	.632**	.362*	.354*	.360*	.520**	0.141	0.127	0.137
2009	Mean PISA reading literacy	.649**	.627**	.591**	.616**	.509**	0.272	0.194	0.250
	90/10 decile ratio (literacy)	-0.105	-0.055	-0.018	-0.037	-.472**	-0.006	0.139	0.067
	Educational deprivation	-.812**	-.794**	-.775**	-.792**	-0.243	-0.183	-0.239	-0.223
	Standard error range (literacy)	0.020	0.057	0.089	0.072	0.118	-0.186	-0.137	-0.174
	Aspired employment at 30 (ISEI scale)	...	...	...	...	...	...	...	...
	Aspired employment at 30 (ISEI scale - STDEV)	...	...	...	...	...	...	...	...

Note: Analysis of two-tailed Pearson correlations using pooled PISA data for four waves in 2000, 2003, 2006 and 2009. Enrolment and Dropout rate figures are calculated using data taken from the OECD Education database (2011). ** denotes significances at the $p < 0.01$ level, * denotes significances at the $p < 0.05$ level. Dropout rates are reversed.

Source: OECD calculations, OECD PISA (various waves), and OECD Education at a Glance (2011).

- On the basis of these results, it is recommended that when reporting statistics using school data, tests for enrolment biases should be performed and countries with significantly lower than average enrolments should be flagged for cautious interpretation.

3.4.1.2 Indicators and relationships with school starting age

180. Across the OECD and Europe, children start school at one of three different ages: 5, 6 or 7 (children in Australia, Israel, the Netherlands, New Zealand and the United Kingdom start at 5, children in Bulgaria, Denmark, Estonia, Finland, Hungary, Latvia, Lithuania, Poland, Romania, Sweden and Switzerland start at 7, in all other countries children start school at 6 – a policy which has not changed since PISA began). The length of time children have been in the school environment may change the extent to which they have become accustomed to teaching practises, including examinations, the school environment and their peers, and as such may better represent capabilities and experiences accumulated by the child than age itself.

181. Figure 2 below compares the outcomes for groups of countries with the same starting ages across five child well-being measures. Charts on the left hand side of the figure compare results using all countries; charts on the right hand side compare results for countries with enrolment rates above 90%. Black squares mark the mean values, blue shaded areas show the error margins, and the whiskers map the standard deviation values. Where white squares are used, this represents comparisons with significant differences in between group variance.

182. To interpret the chart, the case of reading literacy means in 2000 is explained as follows: the black square for the 7-year-old group shows the mean for countries with later starters is around 480, compared to 490 for the 6-year olds starters and just over 500 for the 5 year-old starters. In the same year, for the same indicator, 6 year-olds have the smallest average error margin, showing that the mean value is the most reliable of the three age groups, and the shortest standard deviation bar, showing countries' mean values for 6 year-old starters cluster more closely around the average in that group. Because square for mean reading literacy difference in 2000 remain black there is no evidence (from analysis of variance tests) that the distribution of countries within groups are significantly different.

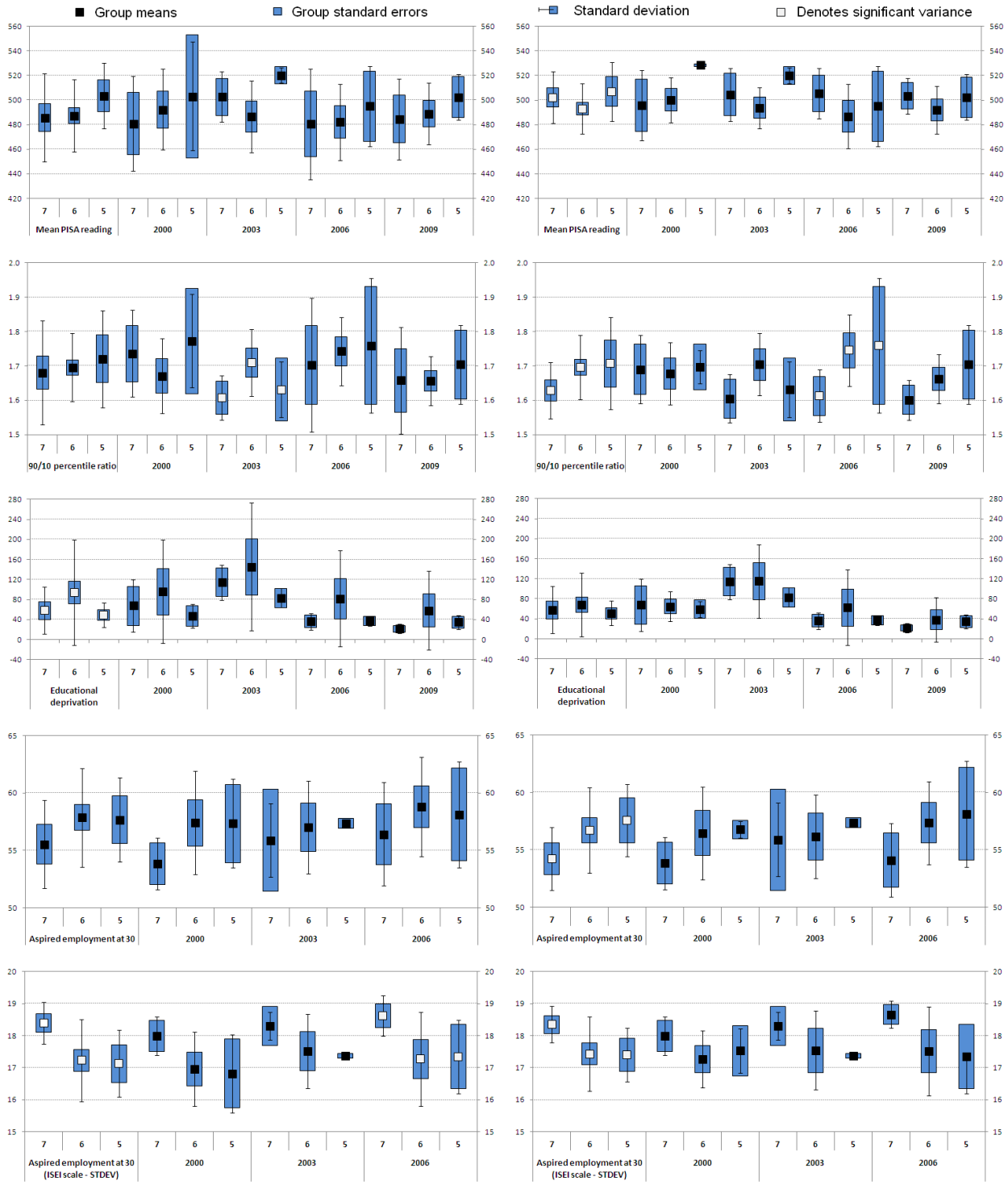
183. Results for the reading literacy indicators (mean scores and 90^{th} percentile over 10^{th} percentile) do not show consistent patterns of years-of-school related gains across the waves. Small, significantly different, results are only found between ages six and seven in literacy distribution scores in 2003 (before sensitivity) and 2006 and overall (after sensitivity) and significant differences between ages five and six in means scores in 2003 (before and after sensitivity) and 2000 after exclusions. For mean values, significant differences always favour the younger starters; in terms of equality in reading scores, results favour older starters.

184. Theoretically, educational deprivation is unlikely to be affected by the years of schooling a child has already received, unless shorter compulsory periods means more available resources for school to provide some items directly to children, or later starts impact on parental work opportunities and so household resources. Pre-sensitivity results show that children starting school at six have higher, but not significantly higher, deprivation. These results are quite clearly driven by low enrolment countries, as deprivation results are much more closely aligned in the sensitized tests.

185. The length of time children spent in formal schooling may affect their employment aspirations either via a longer period in a formal setting or via home influences as parents return to work earlier. Results show that before and after sensitivity tests, average levels of ambitions are almost always slightly higher in school systems with children starting school earlier. Results are not significant until low enrolment countries are excluded and then only for differences between seven-year-old starters and others in pooled tests, and seven-year-olds starters and five-year-olds in 2000. Results for the spread of work ambitions are very similar, with a wider range of choice being expressed in later starting systems. Results for seven-year-old starters are significantly higher than in other systems, before and after sensitivity tests, in pooled tests and in the 2006 wave.

Figure 2: There is little evidence to suggest school starting ages systematically bias child well-being measures

Comparisons of means and variance in child well-being outcomes in PISA by school starting age



Source: OECD calculations using OECD PISA data, and UNESCO education statistics (2011).

Note: Standard errors have been multiplied by 1.96. Significances in variance reported at levels lower than $p < 0.05$.

3.4.2 Sampling and methodological bias

186. Sampling and methodological bias refers to bias arising from the selection of sampling methods, response rates for schools and children (overall and by item), and questionnaire methodology such as the structure and order of items. Child populations vary massively across OECD countries and so sample sizes which rely on absolute numbers of children will affect comparability, statistical reliance on weighting methods, and techniques applied to infer confidence in population estimates. How and when questionnaires asked of children is likely to affect who answers these questions, affecting the accuracy and comparability of estimates country to country, survey wave by survey wave.

187. Most surveys of children apply response rate thresholds for 'primary' and 'secondary' sampling, of schools and pupils, to ensure that sampling frames are replicated, and data collected reflects the sample design and is comparable across countries. However the use of arbitrary response rate thresholds – that is for example 70, 80 or 90 percent of children or schools are required to respond for data to be included in international comparisons – ignores the fact that non-respondents may be significantly different from respondents, that non-response might be non-random, and therefore there is a danger that non-response imposes bias in the national results.

188. Micklewright *et al* (2010) explores the non-response bias in PISA 2000 and 2003 sample for England. The authors show clear evidence that upward bias exists in both samples for the mean score even though 2000 data met the international inclusion criteria and 2003 data were rejected due to low response rates (below response rate thresholds of 80% of schools and pupils after replacement sampling). The authors also find evidence of downward bias in the standard deviation, suggesting higher levels of equity in educational outcomes than in the population (2010, p16). From this evidence the authors argue that there are serious limitations in using arbitrary response rates thresholds as a guide to quality.

189. Non-response is less of an issue – even when response rates are low – where there are no significant differences between populations of respondents and non-respondents (assuming samples are correctly designed). However, as the Micklewright *et al* point out (*ibid*), it is often difficult to test for non-response bias due to the lack of data available for individuals who do not respond to the survey. To overcome these problems researchers can try to match data up with administrative data or - as below - test for non-response to specific items within the survey itself, or cumulative non-response by respondent across the questionnaires.

190. The analysis below will test for significant associations between sample sizes, participation rates in school surveys and key well-being indicators, it will tests for evidence of response attrition and bias in response attrition in background questionnaires, and test for non-response by key child well-being indicators. Where relevant, it will propose basic adjustments to weighting systems to account post-hoc for significant sub-population non-response.

3.4.2.1 Associations between sampling rates / populations / and participation rates and CWB indicators

191. In order to assess the extent to which population sizes, sample sizes and final participation rates might affect the final results reported in any given survey, surveys data over waves is pooled in time series with information on the target population derived from population databases.

Table 11: Higher participation rates are consistently associated to lower error and lower deprivation, but a higher range of aspirations

Pooled time series associations (2000, 2003, 2006 and 2009) before and after sensitivity tests for PISA well-being indicators, population sizes at 15, PISA sample sizes and participation rates

	All countries			Countries with enrolement rates over 90%		
	Population at 15	Samples size	Participation rate	Population at 15	Samples size	Participation rate
Mean PISA reading literacy	-0.136	-0.141	-0.016	0.047	-0.042	-0.051
90/10 decile ratio	0.068	0.000	0.021	0.073	-0.022	-0.023
Educational deprivation	.427**	.247**	-.222*	0.167	-0.121	-.230*
Standard error range	.237**	-.287**	-.452**	.334**	-.327**	-.428**
Aspired employment at 30 (ISEI scale)	.380**	.303**	-0.038	.283*	0.188	0.040
Aspired employment at 30 (ISEI scale - STDEV)	-.308**	-0.048	.330**	-0.175	-0.005	.378**

Note: Analysis of two-tailed Pearson correlations using pooled PISA data for four waves in 2000, 2003, 2006 and 2009. Population figures are calculated using data taken from the OECD Population database (2011), samples are unweighted and taken from PISA datasets. ** denotes significances at the $p < 0.01$ level, * denotes significances at the $p < 0.05$ level.

Source: OECD calculations, OECD PISA (various waves), and OECD Population database (2011).

192. Table 11 presents the pooled time series correlations. In PISA, population sizes, sample sizes, and to a lesser degree participation rates are closely associated to non-literacy outcomes derived from the survey (the outcomes are the same as used in the enrolments analysis above) and the error margins of the reading literacy rates. Following exclusions for the sensitivity tests (again lower enrolment countries) larger populations have marginally significant higher aspirations. Higher participation is weakly associated with lower deprivation, and higher participation is associated to the range of aspired work outcomes.

193. Table 12 shows the same analysis by waves. Besides the error estimates, which again show the clearest association with population and sample estimates (as would be expected), pre-sensitised correlates show few clear patterns, with the exception of the association between educational deprivation and population size. Once countries with low enrolments are removed from the analysis, these effects disappear – suggesting that the role of Mexico in the former association is large and generate false positive results in that test.

Table 12: Outside of reading literacy error margins there is little evidence to suggest systematic sample bias in the child well-being results

Associations, before and after sensitivity tests, for PISA well-being indicators, population sizes at 15, PISA sample sizes and participation rates, 2000, 2003, 2006 and 2009

		All countries			Countries with enrolement rates over 90%		
		Population 15	Sample size	Participation rate	Population 15	Sample size	Participation rate
2000	Mean PISA reading literacy	-0.187	0.265	0.094	0.038	0.309	0.074
	90/10 decile ratio (literacy)	0.050	-0.183	-0.072	0.064	-0.234	-0.132
	Educational deprivation	.464*	-0.062	-0.225	0.294	-0.049	-0.288
	Standard error range (literacy)	.531**	-0.292	-0.328	.596**	-0.300	-0.348
	Aspired employment at 30 (ISEI scale)	.507**	0.207	-0.091	0.391	0.283	-0.036
	Aspired employment at 30 (ISEI scale - STDEV)	-0.328	-0.053	0.263	-0.144	-0.129	0.249
2003	Mean PISA reading literacy	-0.325	-0.289	-0.027	-0.078	-0.083	-0.054
	90/10 decile ratio (literacy)	0.345	0.170	0.056	0.218	0.117	0.000
	Educational deprivation	.378*	.380*	-0.267	0.061	-0.047	-0.267
	Standard error range (literacy)	0.343	-0.033	-.497**	0.264	-0.083	-.494*
	Aspired employment at 30 (ISEI scale)	.527*	0.426	-0.004	.505*	0.129	0.059
	Aspired employment at 30 (ISEI scale - STDEV)	-0.026	0.060	.595**	-0.093	-0.095	.610*
2006	Mean PISA reading literacy	-0.155	-0.140	-0.013	0.069	-0.172	-0.029
	90/10 decile ratio (literacy)	0.056	0.004	-0.073	0.085	0.028	-0.130
	Educational deprivation	.614**	0.280	-0.218	0.107	-0.131	-0.200
	Standard error range (literacy)	0.050	-.339*	-.533**	0.193	-.479*	-.387*
	Aspired employment at 30 (ISEI scale)	0.307	0.282	-0.012	-0.012	0.107	0.135
	Aspired employment at 30 (ISEI scale - STDEV)	-.579**	-0.131	0.289	-.402*	0.068	.418*
2009	Mean PISA reading literacy	-0.024	-0.172	-0.069	0.123	-0.121	-0.165
	90/10 decile ratio (literacy)	-0.030	-0.043	0.207	0.037	0.017	0.201
	Educational deprivation	.493**	.408*	-0.225	0.304	-0.089	-0.244
	Standard error range (literacy)	0.071	-.374*	-.515**	0.259	-0.338	-.560**
	Aspired employment at 30 (ISEI scale)	...	...	...	...	...	...
	Aspired employment at 30 (ISEI scale - STDEV)	...	...	...	...	...	...

Note: Analysis of two-tailed Pearson correlations using pooled PISA data for four waves in 2000, 2003, 2006 and 2009. Population figures are calculated using data taken from the OECD Population database (2011), samples are unweighted and taken from PISA datasets. ** denotes significances at the $p < 0.01$ level, * denotes significances at the $p < 0.05$ level. Countries with lower than 90% enrolment rates were removed from the pooled tests including: Mexico and Turkey in all waves, Portugal in 2003 and 2006 and Luxembourg in 2006.

Source: OECD calculations, OECD PISA (various waves), and OECD Population database (2011).

194. There is little evidence to suggest that there is system-based bias in the child well-being indicators results based on population size, sample sizes and overall participation rates in PISA results. However, the potential for misreading associations skewed by "high population low enrolment" countries (such as Mexico and Turkey) is a cause for concern.

3.4.2.2 Organisation of items in the questionnaires and the effects of attrition

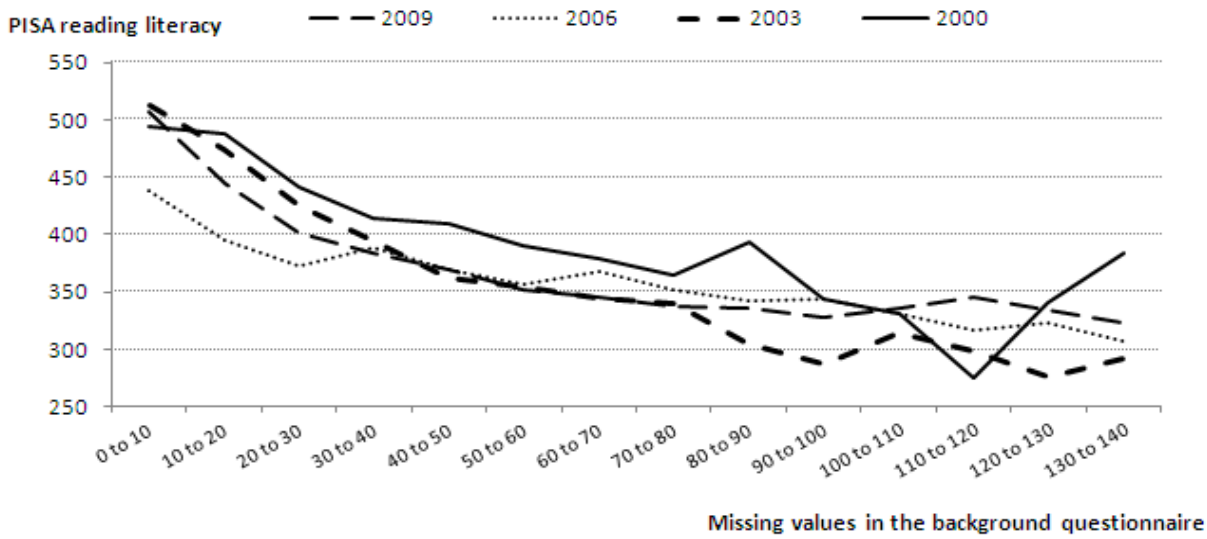
195. In educational assessment surveys, data derived from background questionnaires are essential for interpreting the variation in the achievement indicators, and for developing additional indicators such as various cultural possessions, home wealth, and family form. How a questionnaire is organised is likely to affect the success in collecting the data, for instance later questions are less likely to be answered than earlier questions (as some children lose interest or patience with the task). The important question is whether bias exists in response attrition in questionnaires delivered to children, and whether these vary country to country (or indeed, in comparison to adult surveys).

196. Figure 3 presents the combined OECD EU country data for the four PISA waves. Trends are shown as oppose the correlations for reasons of parsimony. A variable recording the total number of missing values per child in each wave is used to calculate average reading literacy scores for groups of children in categories of missing values by ten (missing values are capped at 140, the total number in 2003 – other waves had more background items). In each wave there is a clear downward trend, showing that the group of children with the fewest missing values have significantly higher PISA reading achievement. This result clearly flags a concern for researchers and policy makers using PISA background data, either to

model variation or to indicate general (student) population estimates on any given background item. There is a concern for coordinators also, particularly in terms of where to place the most important explanatory factors when designing questionnaires.

Figure 3: Less able students are not completing important contextual information for analysis in PISA

The association between Reading literacy in PISA and numbers of missing values in the background question



Source: OECD calculations, OECD PISA (various waves), and OECD Population database (2011).

197. Tests and adjustments for this type of response attrition are necessary by country (see annex chart A.1 for response attrition trends by country). The trends presented by country can be interpreted in terms of flatness and error margins. The flatter the trends of achievement mapped to missing values, the lower the upward educational bias in complex analysis using multiple items from the background questionnaire. The smaller the error margins around the achievement estimates, the less likely random selection of background factors in complex analysis will have on changes in the extent of upward educational bias. In 2009, the flattest trends are seen in Canada, Greece, Hungary, and the United States; the steepest trends are seen in Belgium, France, Japan and Luxembourg. High levels of error around estimates are seen in smaller countries: Iceland, Luxembourg and Slovenia.

198. Figure 3 does not map where in the questionnaire the missing values are, which is also likely to affect final response rates. To do this two tests are undertaken, the first looks at missing values for some key child well-being measures to see whether some items are more sensitive to non-response than others. The second tests whether the position of an item in a questionnaire imposes non-response bias using aspirations data and parental qualifications data; in both cases these data are collected using open-ended questions, questions which have not changed over the waves.

3.4.2.3 Ability in reading literacy and response to selected well-being measures

199. Table 13 compares literacy scores of respondents and non-respondents for nine items from the PISA 2009 survey. The two purposes of this analysis are to, 1) identify the potential for bias in indicators derived from these items, and 2) assess the potential for bias in results when using these items in bivaraite tests and regression equations on literacy outcomes.

200. On each of the items, children with missing values perform worse on average in terms of PISA reading literacy. With the exception of books in the home and whether the mother lives at home the variances in the results are also significantly different (with the exception of repeating grades, the distribution around the mean is larger than for successful respondents). Mean differences are higher in some factors than others. Respondents with missing values on repeated grade, books in the home, preschool attendance, working well in lessons and country of birth are on average around 100 points further down the PISA scale (or broadly one standard deviation) than those with responses. The smallest differences are seen in the fathers and mothers education levels, and father lives at home; these are also the indicators which have the highest rates of missing values (see Annex 6 for a breakdown of missing items by sex and migrant status).

Table 13: Significant differences in means and variances between missing and non-missing values in selected PISA 2009 indicators

	Reading literacy average			
	Missing	Not missing	Significance (t)	Significance (F)
Father's education level	425.27	468.01	Yes	Yes
Mother's education level	404.85	467.56	Yes	Yes
ISCED 2 repeated	378.92	471.23	Yes	Yes
How many books at home?	370.74	466.49	Yes	No
Father lives at home	409.71	468.79	Yes	Yes
Mother lives at home	387.57	467.62	Yes	No
Went to preschool	360.61	466.75	Yes	Yes
Cannot work well in lessons	343.90	467.45	Yes	Yes
Country of birth	360.67	466.40	Yes	Yes

Note: Significance (t) column reports a significant difference in the mean values at the level of $p<0.001$. Significance (F) reports significant differences in the distributions around the mean value at the level of $p < 0.001$.

Source: Secretariat's calculations of OECD PISA 2009.

201. It is clear from this evidence that children who are more literate are more likely to provide useable responses to the items. Moreover, and in the case of indicators commonly used in advanced regression analysis of the datasets, this exclusion leads to upward literacy bias in analysis designed to explain variation in the PISA test scores.

202. Ways to overcome this form of bias could include adjusted weights (see Chapter 6), or adding including the missing values in the analyses. For instance parental education items could be included in the form of a nominal binary matrix designed to capture variations from a reference category (such as 'parent has higher education') which accounts for the missing category.

3.4.2.4 Changes in the placement of the item in a questionnaire and non-response

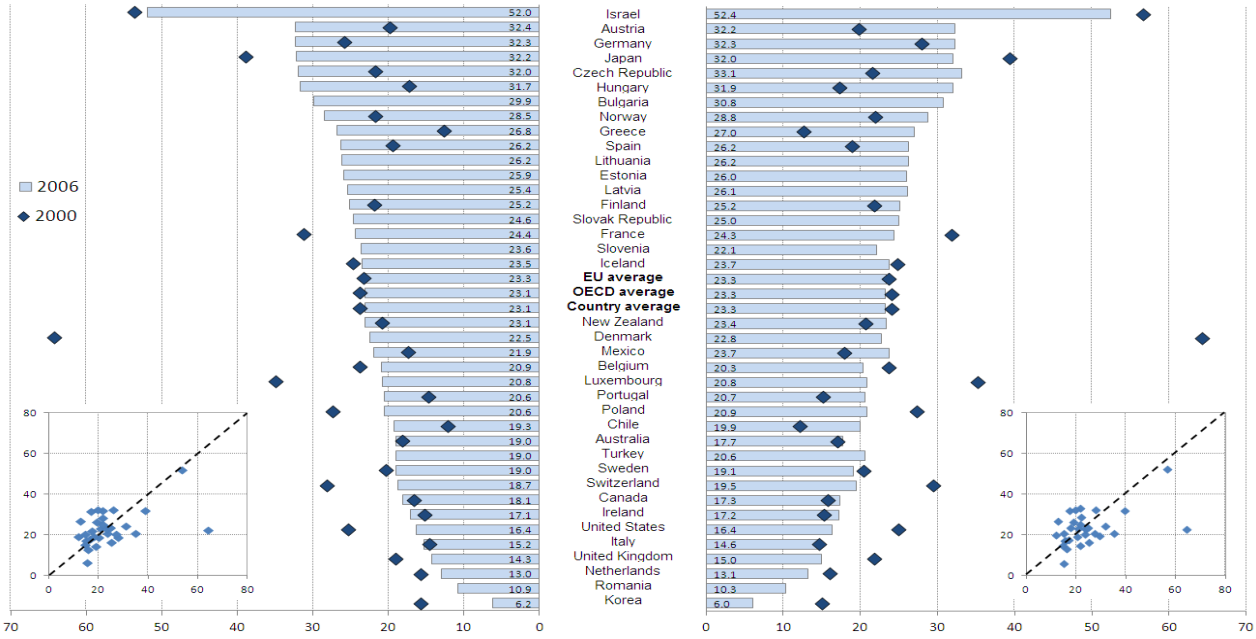
203. The second test looks at the question which asks children about their job aspirations in PISA, which was repeated in the first three waves of PISA (2000, 2003 and 2006), but was dropped in 2009. In 2000 and 2006 the question was included in the main student background questionnaire at item 40 of 42 and at item 30 of 37 respectively. In 2003, the item was placed in a non-compulsory questionnaire (the educational career questionnaire) at item 8 of 8. How have these shifts affected response rates by country?

204. Figure 4 presents the differences in weighted and unweighted missing values for the work aspiration item between the PISA 2000 wave and the 2006 wave.[15] Although non-response is generally high (unsurprisingly given it is at the end of the questionnaire and is an open-ended question), there is little evidence to suggest systematic or excessive unsystematic country-level bias in the results brought on by changes in the placement of the item. Most countries cluster around the 20% missing response rate mark before and after the change in questionnaire placement.

205. Although there is no evidence for generalised differences, two countries stand out as having particular issues. In Israel the high level of non-response on this item is likely to be biased but reliable, and although the change in position is not changing this outcome, the reasons behind this high level of non-response, and potential biases therein, need checking. In Denmark non-response in 2000 stands out strongly from the cluster of countries in the cross plots of the data (see insets). This would suggest that there is a problem with the accuracy of estimates of employment aspirations derived from Danish PISA data in 2000, which is likely to affect the reliability in the trends on this measure for Denmark, and comparability to other countries in 2000.

Figure 4: There is no clear pattern of changing non-response rates in aspirations between the waves

Weighted and unweighted non-response rates for job aspirations questions in the 2000 and 2006 PISA background questionnaires



Source: OECD calculations, OECD PISA (various waves).

206. Changes in the ordering of questions might not only affect the response-rates of indicators but such changes may also affect the accuracy of responses, particularly for more difficult items. What mothers do as their main job is asked in all waves of PISA at the beginning of the student background questionnaires (in contrast to employment aspirations which comes late-on in the questionnaires). In 2000

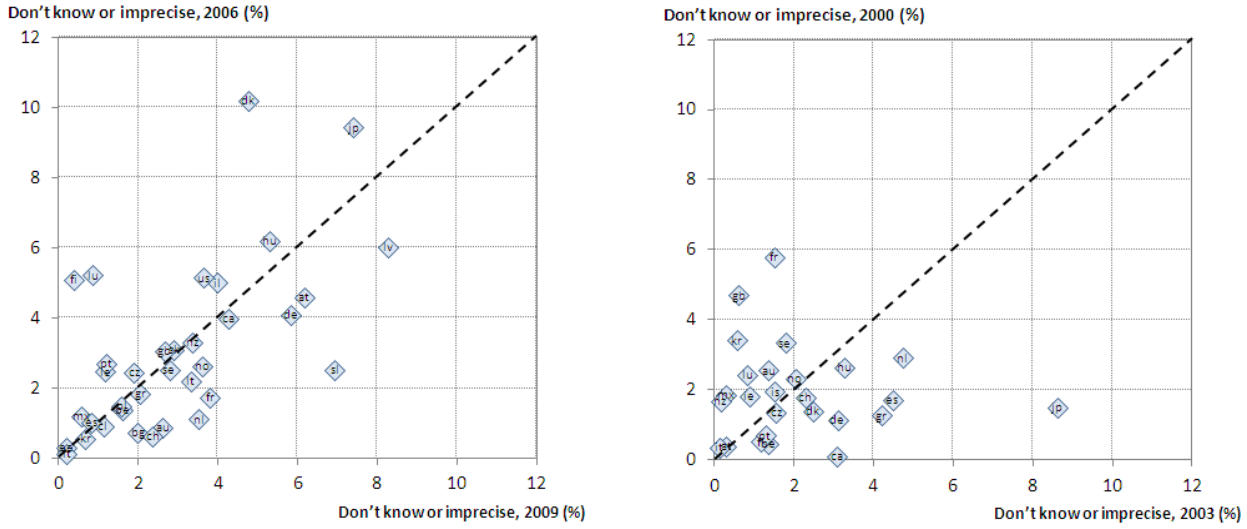
15 In 2003 the inclusion of the employment aspiration question in a non-compulsory questionnaire led to a number of countries not responding at all (these included: Canada, Denmark, Finland, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, and Turkey). Therefore 2003 data is not included in the analysis.

and 2003, students were asked about their mother's job at question number 8, in 2006 at number 5, and at question 9 in 2009. Because it is a difficult open-ended question for many children – the items asks children to 'Please use a sentence to describe the kind of work she does in that job.' – it would be useful to know whether the earlier position in the questionnaire significantly changes the accuracy in responses derived from children.

207. In all waves of PISA, when coding results from the parental employment questions, analysts have the option of recording children's response of 'don't know' and children's responses that are imprecise or vague (for example when children write a good job, or a well-paid job, or a busy job, etc.). By correlating the rates of children who provided imprecise response in 2006 and 2009 (where the gaps in question order was at its largest) and comparing this results to differences between 2000 and 2003 (when the question position does not change) the effect of item order on precision can be explored.

208. Figure 5 presents the result of the test. Overall there is no clear evidence of system-based bias in the precision of results due to the placement of items in the questionnaire. With the exception of Finland Luxembourg, Slovenia and Denmark, changes to the rate of imprecise responses are small. Of more surprise is the results from the 2000 and 2003 test, showing that identically placed questions can produce big changes in rate of imprecise answers between the two waves.

Figure 5: Evidence suggest that the placement of complex items in the background questions do not systematically affect the precision of responses on parental employment



Source: OECD calculations, OECD PISA (various waves).

3.4.2.5 Indicators non-response bias in PISA

209. An important question when trying to assess the accuracy of an indicator for representing the population is whether non-respondents are different from the people representing them as respondents. This section briefly presents multivariate analysis of non-respondent groups for four key indicators selected from the four different waves of PISA. These indicators include: parental relationships (parents discussing school problems with children) in 2000, maternal employment in 2003, students' employment aspirations in 2006, and educational deprivation in 2009. Results are presented in annex tables A.2 through to A.5.

210. What is evidently clear from the results is that significant non-response bias is likely in many of the key indicators derived from the surveys of children and that adjustment weights should be applied to

take account of non-response bias as much as possible (see 5.5.2 for an example and caveats). These adjustments weights will have to be applied on a country by country basis, and on an indicator-by-indicator basis depending on the population appearing as significant non-responses groups, in different countries across the different indicators.

211. In 2000, weighted data for parental interaction produced only 11 examples of countries with significant non-response by sex (despite the significance being calculated using inflated N's), migrant status or parental education. The level of significant non-response ranged from small (non-responding boys in Canada) to large (migrants in Hungary). In 2003 (mothers employment), 2006 (employment aspirations) and 2009 (educational deprivation) significances are so common, and mid-range in size – with the exception of educational deprivation – there may be the case for preparing numerator values for complete sample data, as well as standard calculation methods (even syntax models) for the 'easy-to-use' adjustment weights by country specifically for the purpose of indicator development in these waves.

3.4.2.5 Indicators non-response bias in HBSC

212. Over the past three waves of the HBSC survey, students' rates of response have decreased cross-nationally. In the 2001/2002 survey, the percentage of unanswered questions ranged from six percent in Estonia (where the average student missed seven questions) to 20 percent in Austria (where the average student missed 24 questions out of 121). By the 2009/2010 wave, the percentage of unanswered questions ranged from eight percent in Portugal (where the average student missed ten out of 125 questions) to 24 percent in Turkey (where the average student missed 31 questions, many of which were excluded by national coordinators of the survey). Between 2005/2006 and 2009/2010, the middle of the distribution also rose indicating that, on average, most countries experienced increases in rates of non-response.[16] These per capita rates of response vary considerably across countries over time, and bring with them a need to test the validity and reliability, both within and between countries, over survey waves.

213. Notably, however, lower rates of response cluster around specific survey questions important for assessing child well-being. For instance, in the 2001 survey, nearly nine percent of students did not answer the question "How much do you weigh without clothes?"; 11.3 percent of students did not answer the question "How tall are you without shoes?"; and eight percent of students did not evaluate their "life satisfaction." In contrast, students nearly universally provided standard background information such as gender and age, and fewer than one percent of students abstained from questions on bullying and enjoyment of school.

214. Having identified life satisfaction, weight, and height as examples of commonly unanswered questions, the results of simple tests for bias in non-response to these questions – across countries and waves – are introduced below (for tables detailed results for all countries, see Annex 9).

3.4.2.5.1 Frequency and correlates of non-response to the "life satisfaction" query

215. What factors make some students more likely than others to answer questions about their positive health? Specifically, why do some students choose not to describe their life satisfaction? Four potential correlates of non-response are evaluated here: sex, age, wealth, and health.[17] These four independent

[16] All results in this section are based on Author's calculations of raw data for the last three waves of the survey provided by HBSC network.

[17] Wealth is measured using the responses to "How well off do you think your family is: very well off / quite well off / average / not very well off / not at all well off?". Health is measured using responses to "Would

variables are evaluated in a binary logistic model in which the outcome is a dummy (0/1) variable for the likelihood of response to the life satisfaction question (1). The statistically significant results of this model are presented in Table A9.1 of Annex 9. After running the logit model for every EU27 or OECD country in the survey waves 2001/02, 2005/06, 2009/10, the number of statistically significant correlations between each of the independent variables and the dependent variable across the three surveys has been tabulated.

216. To interpret Table A9.1, consider the case of Belgium. In one of the three surveys, age is statistically significantly correlated with the likelihood of responding to the "life satisfaction" question, with younger children less likely to respond, and older children more likely to respond. Respondent's sex has no significant correlation with likelihood of non-response to the life satisfaction question. Lower wealth predicts the likelihood of response in all three of the HBSC surveys in Belgium, and health is significantly associated with the likelihood of response in two waves of the survey, when healthier students were more likely to answer.

217. Looking across the all countries with cases of response bias, age and gender are the only factors for which the "sign" or "direction" of the correlation is unidirectional: younger students respond less often than older students, and boys respond less often than girls, in every statistically significant correlation.

218. In most countries with health-induced response bias, students who self-describe as having poor health are less likely to answer questions about their life satisfaction (Denmark is the exception in which healthier students are less likely to answer this question).

219. Interestingly, the direction of the correlation between wealth and response outcomes is somewhat unpredictable. Wealth is a common correlate of likelihood of response, but the direction of the correlation varies across countries and over time. In Lithuania, the Netherlands, Turkey, and for two years in Canada, wealthier students are more likely to respond; in all other countries, low-income students are more likely to respond. Consequently, common assumptions about the socioeconomic status of students who underreport life satisfaction across countries cannot be drawn.

3.4.2.5.2 Frequency and correlates of non-response to the 'weight' and 'height' query

220. Two important indicators in the HBSC survey are those that record self-reported height and weight. Height and weight are important indicators as together they contribute to the only internationally comparative measures of BMI in childhood, and they are therefore important indicators of pre-obesity and obesity. Moreover, the ways in which these data are collected in different countries change due to the differences in the standard measures used (metric or imperial); this means that cultural bias may creep into comparisons on this measure. Finally, obesity in childhood has increasingly come to the attention of policymakers in a number of countries (such as the U.S. and the U.K.), and so reliable and representative measures of BMI are needed to analyze and inform child policy development across OECD countries.

221. What factors influence whether or not a student provides a response to the question about their weight? The four likely correlates of non-response used in the life satisfaction test are again evaluated here: sex, age, wealth, and health. These four variables are evaluated in a binary logistic model in which the outcome is a binary variable for the likelihood of response to the weight question (1).

---

you say your health is excellent / good / fair / poor?". These measures are reverse-coded with higher numbers representing poorer outcomes.

222. The main results of this model are presented in Table A9.2 of Annex 9, where the number of statistically significant correlations is tabulated between each of the independent variables and the dependent variable across the three surveys.

223. Across countries, a student's sex is a frequent and significant predictor of whether or not the student answers the weight question. In nearly all of the countries with response bias in the weight question, boys are more likely than girls to provide their weight. Only in Lithuania and Estonia are girls more likely than boys to provide their weight.

224. As with life satisfaction, age is a consistent and positive predictor of response. That is, older students are more likely to answer (and younger students are less likely to answer) questions about their weight. In terms of health, students who self-report as "more healthy" are consistently more likely to answer the weight question. The exceptions to this are the United States and Germany, where self-reported "less healthy" students are more likely to answer the weight question.

225. Again, the direction of the relationship between wealth and likelihood of response to the weight question is unpredictable – although only eight countries report bias. Low-income students are more likely to respond in half of the countries: Belgium, Germany, Latvia, and the U.S. High-income students are more likely to answer the weight question in the other half of countries: the Czech Republic, Greece, Lithuania, and the Netherlands.

226. Moving on to height, what factors influence whether or not a student provides their height to HBSC? This question is presented adjacent to weight and is asked early-on in the survey; weight is the thirteenth question and height is the fourteenth question (out of over 100) in the HBSC surveys. The same four likely correlates of non-response are again evaluated here: sex, age, wealth, and health. Correlates are evaluated in a binary logistic model in which the outcome is the likelihood of response to the height question. The results are presented in Table A9.3 of Annex 9 using the methods described above.

227. With the exception of Portugal, every country experiences age-related response bias in the height question in at least one survey wave. In every country-year combination, the correlation between age and response is positive: older students are more likely to provide responses on their height, and younger students are less likely to provide responses on their height.

228. In the majority of countries with gendered response bias, girls are more likely than boys to provide their height. The exceptions are Turkey, Ireland, and Belgium, in which girls are less likely than boys to give their height, and Switzerland, where there are mixed results across surveys. The direction of the relationship between wealth and likelihood of response again varied across countries and time, and so common cross-national conclusions about socioeconomic status and information provision cannot be drawn. Poorer students are more likely to respond in Estonia and Spain (in one survey year), and higher-income students are more likely to respond in Lithuania, Sweden, Latvia, Ireland, the Czech Republic, and Poland.

229. In every country with health-correlated response bias, students who describe themselves as healthy are more likely to provide information on their height.

3.4.3 Cultural and linguistic bias

230. Blum *et al* (2001), in an assessment of the International Adult Literacy Survey, find for the IALS that the premise of validity of cross-national surveys '…based upon a strong hypothesis of an identical difficulty scale of tasks among cultures and languages' (2001, p233) is unsafe. The authors uncover a number of examples in IALS where translations and cultural references put speakers of one language in one country (French speakers and France) at a disadvantage when answering questions in the assessment

which generates the literacy scores. Examples include precision in translation (and repetition), and errors both in translation and not establishing 'culture-free' questions.

231. In comparing French with English translations a lack of precision in translation of questions leads to varied detail of responses. Blum *et al* highlight a question in IALS which asks respondents to 'list all the rates' in English but simply asks 'What rates' in French. This, they argue will mean French respondents answer the question in less detail, in some cases providing a single answer response, when a list was required (Blum *et al*, 2001, p 235). Translations may introduce issues of repetition. For instance when the question and answers use the same term in one translation, but two different terms in another, the latter example requires an extra step of cognition (associating the two terms) before the respondent can fully answer the question (ibid, pp 234-5). Evidence of this type of error was found for different English and French translations in IALS.

232. Errors in the IALS survey due to insensitivity to differences in culture knowledge were also evident. For example, cultural bias could enter the results was a question which included a French actor in an answer to a specific film question (Blum *et al* 2001). Cultural expectations of the actor's film roles are argued to have distracted some French respondents into answering the question incorrectly; in countries where respondents did not have expectation or knowledge of the actor, it is unlikely they would have had to contend with the distraction (ibid).

3.4.3.1 Linguistic bias in translations in PISA

233. Linguistic tests are difficult to undertake on all different language questionnaires simultaneously – and is a barrier to such tests in this project. Using PISA data, another opportunity to test translations comes in the form of options to introduce colloquial terms in the background questionnaires. The example used here is of children being asked to report the number of cellular phones in their home (none, one, two, three or more).

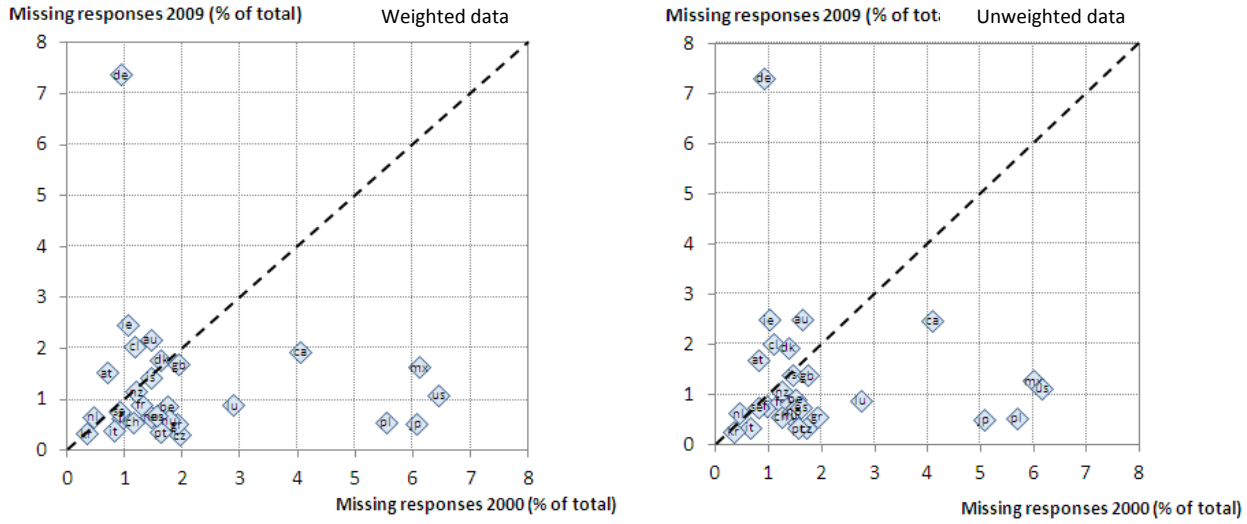
234. Questions 22 and 14 of the PISA background questionnaires in 2000 and 2003 respectively asked children "How many of these do you have at your home?". Cellular phone was listed top and presented in French quotation marks like so: <cellular phone>. In 2006 and 2009 the question had changed slightly to "How many of these are there at your home?" and cellular phone was listed again, but without the option of including a colloquial translation. Given that two waves allowed country to modify the term and two waves did not this provides an opportunity to explore the effect of introducing colloquial terms in translation practises on response rates (bearing in mind that the question has changed from a stress on personal ownership 'do you have' to a stress on family ownership 'are there at your home').[18]

235. Figure 6 below plots 2000 to 2009 weighted and weighted non-response rate for ownership of cellular phones. Although non-response rates are small overall (unsurprising given the simplicity of the question), with more countries falling below the 45 degree line there is broad evidence to suggest that non-response is lower in 2009 in most countries when colloquial translation is no longer available (this may in part be due to more children being able to answer this question positively, as other family members owned this item). In a handful of countries, including Mexico, the United States and Japan, response rates increased substantially. Only in Germany in 2009 have non-response rates increased to any degree.

[18] A second example of changes in colloquial translation is shown for the sex selection question (male / female). In 2000 and 2003 translators and national managers could choose to use boy and girl, this was not explicitly the case n 2006 and 2009.

236. The use of colloquial translation does not seem to have a systematic effect on the non-response rates of children to this item. However, there are a number of countries where results differ strongly, and for these reasons alone, non-response by country should be investigated further and excluded from comparisons and time trends analysis if necessary.

Figure 6: In 2009 fewer German children provide useable responses to the cellular phone ownership question



Source: OECD calculations, OECD PISA (various waves).

3.4.3.2 Reporting place of birth by children and response sensitivity to country-specific options

237. Migrant status, particularly when accounting for the similarities and differences between the host country and the country of origin, is an important factor for identifying well-being risks and resilience. For these reasons PISA changed its question to children about migrant status in 2003 to offer a list of countries where children or their parents were born (this has remained unchanged by 2009 – in 2000 the questionnaire asked simply whether children and parents came from the test country or not). Using this change, it is possible to assess whether children are more or less inclined to complete these questions under the assumption that identifying with specific countries has positive or negative incentive for disclosing birth place or heritage.

238. Table 14 report the results for weighted non-response to the different migrant status questions for each member of the family form PISA 2000 and 2009. This question is always delivered in the first few pages of the PISA student background questionnaire, and suffers from very low non-response rates.

Table 14: Most countries have seen small non-response rates for parents' migrant status since 2000

Weighted non response rates to place of birth questions in PISA in 2000 and 2009

	Migrant (%)		Dif.	Migrant mother (%)		Dif.	Migrant father (%)		Dif
	2000	2009		2000	2009		2000	2009	
Australia	1.35	1.71	.36	.43	.61	.19	.58	1.18	.60
Austria	.76	1.62	.86	.44	.90	.45	.46	.94	.48
Belgium	1.33	1.59	.25	.67	1.16	.49	.54	1.25	.71
Bulgaria		2.29			1.93			1.62	
Canada	1.73	1.84	.11	.57	1.42	.85	.59	1.58	.99
Chile	1.19	2.42	1.23	.72	.86	.15	.86	1.43	.57
Czech Republic	3.32	.53	-2.79	2.96	.58	-2.39	2.74	.66	-2.08
Denmark	1.04	2.33	1.29	.66	1.33	.67	.57	1.16	.60
Estonia		.53			.87			1.29	
Finland	.99	.81	-.18	.68	.79	.11	.60	.64	.04
France	1.07	.86	-.21	.86	.47	-.39	.86	1.02	.17
Germany	1.73	7.87	6.14	1.40	2.27	.87	1.30	2.19	.89
Greece	1.28	.68	-.60	.47	.89	.41	.49	.80	.31
Hungary	.57	.72	.14	.51	.63	.12	.43	.78	.35
Iceland	1.19	1.73	.54	.33	.99	.66	.33	1.01	.69
Ireland	1.30	3.61	2.31	.44	1.83	1.39	.29	1.91	1.62
Israel	11.92	2.92	-9.00	2.18	1.84	-.34	2.36	1.84	-.52
Italy	.72	.52	-.20	.44	.79	.35	.50	.72	.22
Japan	5.04	.39	-4.65	.08	.23	.15	.04	1.07	1.03
Korea (*)		.32			.96			.92	
Latvia		.51			.76			1.40	
Lithuania		1.48			2.36			2.21	
Luxembourg	2.15	1.21	-.94	.68	1.36	.68	.85	1.73	.88
Mexico	3.83	2.08	-1.75	1.26	1.20	-.06	1.15	1.58	.43
Netherlands	.56	1.26	.70	.28	.50	.22	.16	.55	.39
New Zealand	1.85	1.14	-.71	.93	.62	-.30	.90	1.01	.11
Norway	1.45	.71	-.74	.75	.41	-.34	.80	.32	-.47
Poland	1.31	.77	-.54	.71	1.24	.53	.68	1.00	.31
Portugal	1.11	1.14	.03	.35	.67	.32	.37	.71	.34
Romania		.82	.82		4.08	4.08		2.51	2.51
Slovak Republic		.48			.46			.57	
Slovenia		1.27			.73			.78	
Spain	1.42	1.35	-.06	.69	1.09	.39	.68	1.06	.39
Sweden	.82	1.29	.48	.32	.42	.10	.29	.59	.30
Switzerland	1.07	1.14	.08	1.10	1.19	.09	1.05	1.06	.01
Turkey		.78			1.50			1.30	
United Kingdom	1.35	1.87	.52	.63	1.27	.64	.48	1.63	1.15
United States	5.07	1.78	-3.29	.39	1.03	.64	.29	1.68	1.40
Average	**2.02**	**1.65**	**-.37**	**.76**	**.99**	**.23**	**.73**	**1.14**	**.41**

Source: OECD calculations, OECD PISA (various waves).

239. Overall results are mixed, showing that children's self-reported migrant status is on average higher with country specific options, although big shifts are seen towards lower non-response rates in the USA, Japan and Israel, and higher non-response in Germany. For parents' migrant status, only 5 countries for mothers and 3 countries for fathers show improved response rates when country-specific options are available, although in no cases are shifts sufficiently large to suggest any for of significant bias in the results for parents' migrant status.

3.5 Comparisons of interchangeable measures: external and internal reliability

240. This section will review key child well-being indicators that are repeated across the surveys as a final validity and reliability check on the main ranking comparisons used in international comparisons of child well-being. Four comparisons are undertaken. The first will test the validity of the Family Affluence Scale produce by HBSC, by recreating the composite measure using PISA data. The second will assess the validity of distributions produced in educational achievement surveys over time, to infer a certain level of

confidence in distributional reliability, and therefore inequality measures. The third will compare data used for calculating poverty figures derived from household surveys (EU SILC and ESS), and the final section will compare indicators of risk taking from the HBSC and ESPAD.

3.5.1 Family affluence in PISA and HBSC

241. Using the PISA survey for 2000[19] Family Affluence Scale (FAS) composite used in the HBSC surveys and International Reports can be reproduced and tested for validity. This is an important question when considering issues of social gradient, as if affluence measures are reproducible, analysis of the type undertaken in Currie *et al*, 2008) can be complemented with similar social gradient analysis of PISA results.

242. Questions used in FAS regarding the ownership of a family car, a room of your own and number of family computers are all repeatable in PISA 2000 (including the numbers of items of each required to create the composite). An equivalent item for holidays taken by children was missing from PISA. To recreate the affluence aspect of 'holidays' for the PISA FAS results, data on whether children 'Attended a popular music concert', 'Attended an opera, ballet or classical symphony concert', 'Watched live theatre' or 'Attended sporting events' were included instead.[20] Children who had experienced any of these sorts of events are expected also to be children in families with sufficient incomes to take regular holidays (the range of events is included in order to account for tastes).

243. HBSC Family affluence scale is computed as follows:

- Number of cars (1 = 0 points) (2 = 1 point) (3 = 2 points);
- Having own bedroom (No = 0 points) (Yes = 1 point);
- Number of holidays / computers (1 = 0 points) (2 = 1 point) (3 or 4 = 2 points);
- Family Affluence scale = cars + holidays + computers + 1 if child has own room;
- The family affluence scale is then categorized (0, 1, 2, 3 points = Low) (4, 5 points = Medium) (6, 7 points = High).

244. The PISA family affluence scale is computed as follows:

- Number of cars (0 = 0) (1 = 1) (2 or more = 2);
- Having own bedroom (No = 0) (Yes = 1);
- Number of cultural activities (never or hardly ever = 0), (once or twice = 1), (3 times or more = 2);
- Number of computers (0 = 0) (1 = 1) (2 or more = 2);
- Family Affluence scale = cars + cultural activities + computers + 1 if child has own room;
- The PISA family affluence scale is then categorized (0, 1, 2, 3 = Low) (4, 5 = Medium) (6, 7 = High).

245. Table 15 shows the cross national percentages for the 'low' category for each of the family affluence constructs. All countries with available data for both surveys have been included.

---

19 Will include updates correlations for later waves.

20 These choices on these measures were 'Never or hardly ever', 'Once or twice a year, 'Three or four times a year', and 'More than four times a year'. The results were recoded into zero, one and two or more in the order of the above (aggregating last two responses). The highest response across the four activities was recorded for each child.

246. The cross-plot in Figure 7 shows correlation coefficient of r=0.91 (R^2 of 0.82) between the PISA FAS scale and the HBSC findings[21]. The results here provide evidence of external validity in both the PISA family wealth items and the HBSC family wealth variable when composited in the form of a summed scale. This means that researchers can have confidence that the commonly used HBSC FAS scale is externally valid for the 2001/02 wave (tests will need repeating for 2005/06 and 2009/10).

Table 15: Percentage of youth population in each survey reporting low affluence by type of scale

Proportions of children reporting low family affluence on PISA (2000) and HBSC (2001/02) family affluence scales

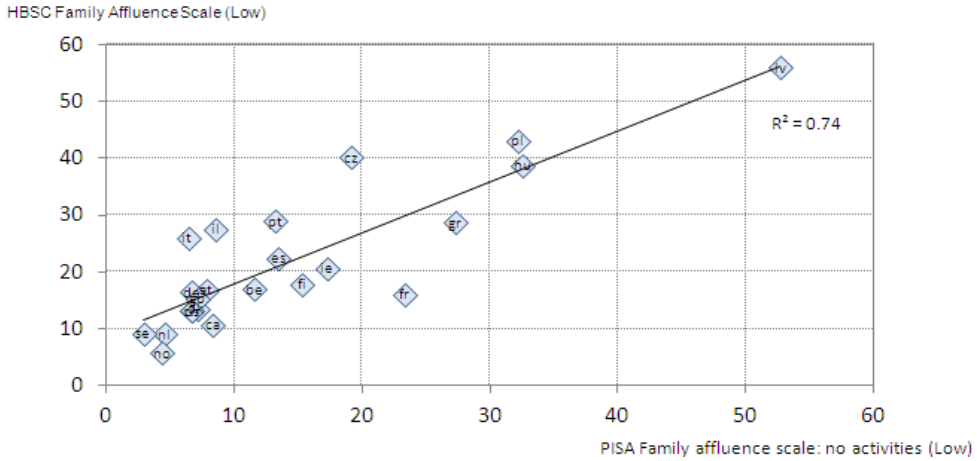
	PISA Family affluence scale (Low)	PISA Family affluence scale: no activities (Low)	HBSC Family Affluence Scale (Low)
Austria	5.9	7.9	16.8
Belgium	10.7	11.6	16.9
Canada	4.5	8.3	10.7
Czech Republic	7.2	19.2	40.2
Denmark	5.0	7.1	13.5
Finland	12.6	15.3	17.8
France	18.8	23.3	16.1
Germany	6.0	6.7	16.4
Greece	12.2	27.3	28.7
Hungary	13.0	32.6	38.7
Ireland	6.6	17.3	20.7
Israel	6.8	8.6	27.5
Italy	5.8	6.4	26.0
Latvia	23.8	52.7	55.9
Netherlands	8.0	4.6	9.0
Norway	4.9	4.4	5.8
Poland	19.7	32.2	43.1
Portugal	5.1	13.2	28.9
Spain	8.2	13.4	22.4
Sweden	3.5	2.9	9.2
Switzerland	4.7	6.7	13.1
United Kingdom	6.2	7.2	15.3
United States	4.0	6.7	13.1
Country average	**8.8**	**14.6**	**22.0**

Note: Children in the HBSC sample are aged 11, 13 and 15. In PISA children are aged 15 years.

Source: OECD PISA 2000 and HBSC 2001/02 (Currie et al, 2004). OECD calculations.

[21] A correlation was also run on the 21 countries included in the OECD Child well-being analysis. The coefficient for this test was only marginally weaker at r=0.88.

Figure 7: Correlation between PISA FAS calculations and HBSC FAS calculations



Source: OECD PISA 2000 and HBSC 2001/02 (Currie et al, 2004). OECD calculations.

3.5.2 Literacy distributions in PIRLS and TIMSS

247. This section will introduce comparisons of achievement indicators distributions across countries undertaking the TIMSS and PIRLS surveys. PISA is not included in this comparison due to differences in sampling and definitions of literacy. The purpose is to test the validity of distributions produced in educational achievement surveys over time, to infer a certain level of confidence in distributional reliability, and therefore the potential for producing robust inequality measures.

248. Distributions for reading, mathematics and science literacy have been produced for TIMSS and PIRLS over all available waves, and selected countries, and are presented in Annex 7 of this document (Figure 1 for TIMSS and Figure 2 for PIRLS). To calculate the distributions, data across all waves for all countries, and both surveys, have been standardised to a mean of 50 and standard deviation of 10.

249. TIMSS data across all waves are normal for all countries, for which there are data, with no significant skew in the distribution, except for in Japan and Korea where the distribution is negatively skewed, while the distribution for Chile and Turkey are positively skewed. For the four countries with significant skews in the distribution it is important to note that the skew persists across all survey waves. This suggests that either the skew is a true reflection of the ability of students in these countries rather than a form of sampling bias (a view supported by the distribution of the measures of mathematical ability as tested by the PISA survey), or a bias has persisted across the survey waves.

250. In many countries the distribution of TIMSS literacy measures have shifted from a slight negative skew in 1995, through the waves, to a slight positive skew in 2007. This can be seen in the figures for countries where the 1995 line starts “lower” than the others (i.e. the median is slightly above 50 - above the mean) and the data in the top half of the distribution is more “bunched up”, suggesting a slight negative skew. This trend in shifting distributions is not universal, and importantly the differences in the skew across the waves are not significant.

251. In contrast to TIMSS, data for PIRLS normal across all waves for all countries, for which data are available, with no visibly large skew in the distribution. The fact that there is less skew in the distribution of PIRLS may be due to more normal distribution of abilities across a given population at a younger age, or a more normal distribution for reading abilities when compared to mathematics abilities

(the TIMSS survey measures the mathematics ability among 8th grade pupils, while the PIRLS survey measures the reading ability across 4th grade pupils).

252. What is notable is that PIRLS distributions for 2001 are less spread (as can be seen by the dotted line being "below" the continuous line in the lower half of the distribution and above the continuous line in the top half of the distribution). The underlying data behind the chart does indeed confirm that the standard deviation for 2001 is a slightly less than 10, raising some concerns regarding the standardisation procedure.

253. Overall, the similarity in the shape of the distribution in both TIMSS and PIRLS indicates that the tests are capturing similar measures of ability across the waves. This allow users to compare results across the waves with a greater degree of reliability and provides greater confidence that any variations observed are a result of changes in ability in the target population rather than a change in the testing methodology. Also, any subsequent measures of the distribution (such as inequality) can also be calculated with greater confidence.

254. However, some caution must be taken when interpreting the results for both surveys. While the similarity in the distribution allows for greater confidence when comparing across waves, it should also be borne in mind that this also suggests that if any large form of bias (such as hidden populations – see 5.5.1) does exist in the sample, it is likely to have persisted across the waves. Also, as with PISA, while the standardisation facilitates certain comparisons across the waves it removes the ability to be able to detect whether the absolute performance of the target population as a whole has improved/worsened across the waves.

3.5.3 Self reported household structure and income in EU SILC and ESS

255. This section will introduce comparisons of self-reported adult data for the calculation of income poverty rates in the household surveys. In particular it will compare self-reported household family structure and income in the European Social Survey (ESS) and the EU Survey on Income and Living Conditions (EU SILC).

256. Basic difference between the methodology of the two surveys (in the survey cycle and in the country coverage) will create certain restrictions for the comparison. For example, the ESS is conducted every two years since 2002 and the last dataset available is for 2010, whereas EU SILC is conducted every year since 2002, and the last available round at the time of writing is for 2009. This means comparisons of the results are undertaken for the last round in common, 2008. Moreover, in 2008 not all countries were common to both surveys (of 31 countries in total, EU SILC covered 28, and ESS 27); mismatches and ESS countries where income data is not available (Bulgaria, Cyprus, Lithuania and Slovakia) meant 21 countries are included in the analysis.[22]

3.5.3.1 Comparing family structures

257. Accurate information on family structure is essential for comparing risks by family type across a range of child well-being measures, as well as for equivalising net family incomes to assess poverty rates.

258. Figure 8 plots the proportion of 3 family types and all families with children as reported in ESS and EU SILC in 2008. Estimates of the proportion of households with children in ESS and EU SILC are

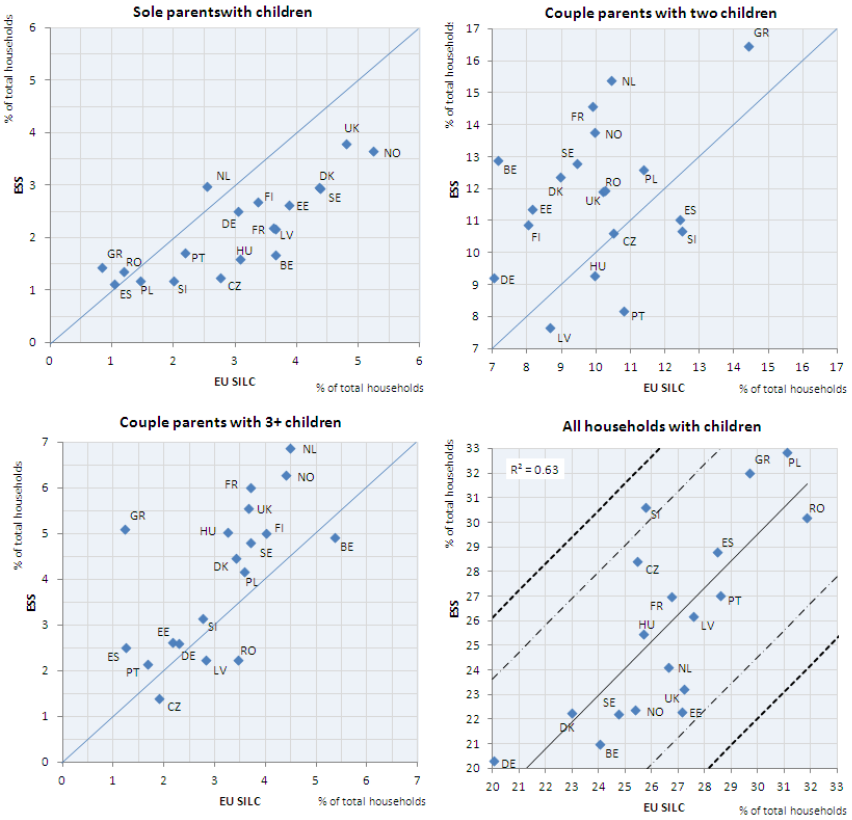
[22] Weights have been applied to the raw data in both surveys to avoid sampling bias. In ESS, the design weights have been used to correct for the slightly different probabilities of household selection. In EU SILC, the household weight has been used to improve the accuracy of the estimates.

very similar (see bottom right-hand chart of Figure 8). The difference is only larger than 2 or 3% in a few countries (Estonia, Slovenia, and the United Kingdom). However, a more detailed breakdown by family type shows larger discrepancies. Sole parents are under-represented in ESS when compared to EU-SILC[23], and couple parents with two children are over represented. In both cases, most of the countries are at the same side of the diagonal line – though notably the trend for sole parents is much more uniform, suggesting reliability if not validity. For larger families (three or more children), ESS over represents in comparison to EU SILC, but unlike couple families, and with the exception of Greece and Belgium, there is some uniformity in the differences.

259. Although results are not encouraging, with the exception of couple families with 2 children, most differences could fall into margins of acceptable error on single indicators (plus 3 or minus 3 percent). Of concern, however, is that combined error by type will provide a very different picture of families depending on which surveys is chosen, and impact on standard income estimates and poverty after equivalisation. Also in cases where populations may be small (such as sole parent families, or large families) analysis of this data as population estimates without adjustments is not recommended – on occasion reliability in the sets may allow for adjustment weights to be calculated, using the larger surveys as the numerator.

Figure 8. Proportion of households with children is similar in ESS and EU SILC

Percentage of households with children in total households, 2008



Source: Author's calculations of ESS 2008 and EU-SILC 2008.

[23] The overwhelming tendency would be to have greater confidence in SILC figures due to the methodological strengths it has over ESS (see section 2.5.3), which explains the order of comparison.

3.5.3.1 Comparing self-reported income in deciles

260. Using the same surveys and waves, income comparisons can be undertaken at the group level. Although EU SILC collects the exact income figures, deciles are the common denominator, given that ESS collects information on income asking households in which national income group they fall. To do this, each household income from EU SILC has been allocated to one of these ranges reported in ESS. The distributions were then fully comparable.

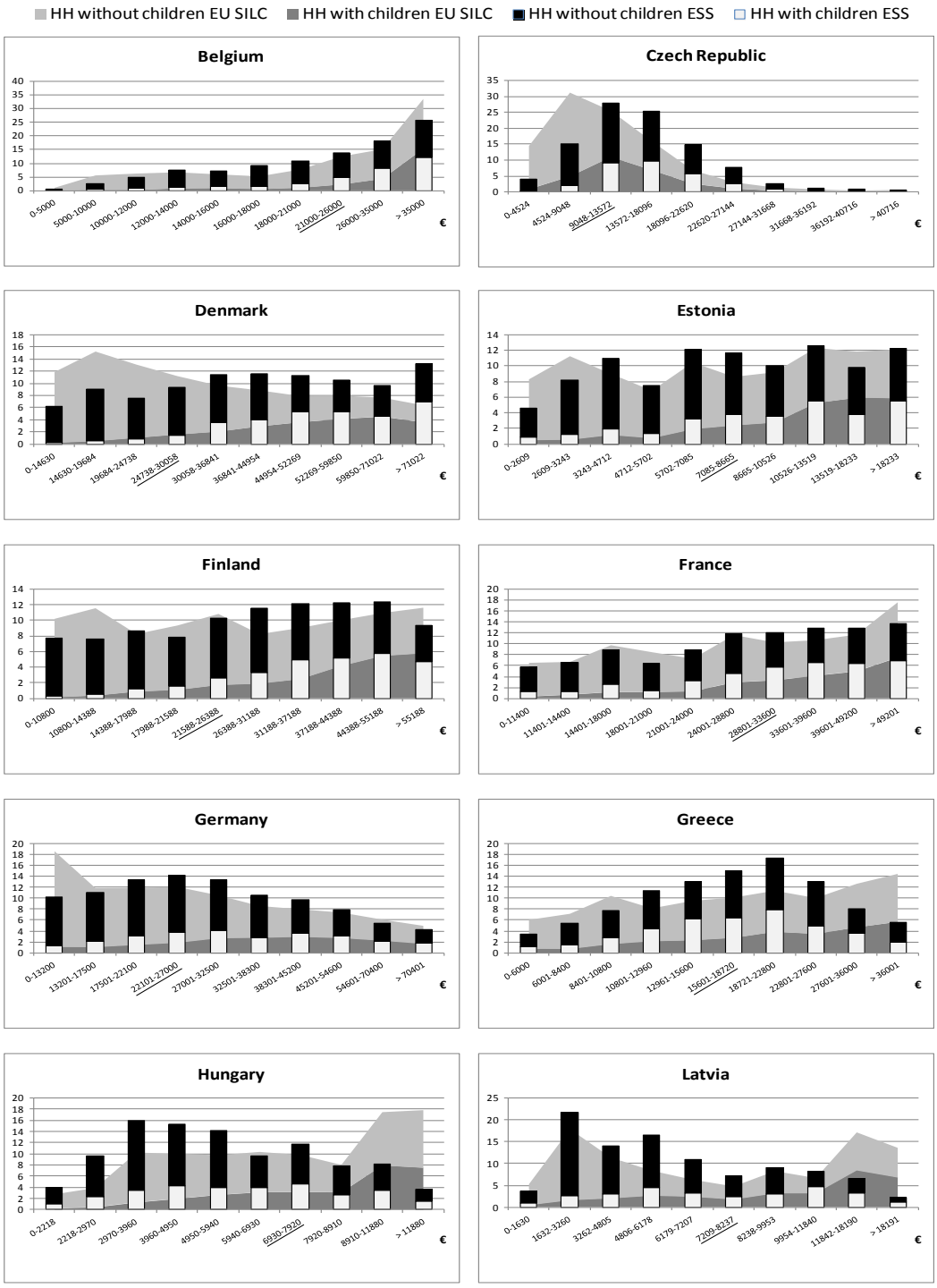
261. Figure 9 contains the distributions of household income in ESS and EU SILC for the 19 countries of the comparison. On each chart, proportions of ESS respondents in each income bracket are represented by the bar chart, with EU SILC proportions in the area behind. Proportions are broken down by households with children and households without children, the reason being that both groups' incomes contribute to the calculation of the poverty threshold.

262. The household income distribution presents similar trends and levels in ESS and in EU SILC in eight countries: Belgium, the Czech Republic, Estonia, France, the Netherlands, Romania, Slovenia and the United Kingdom. For these countries, there are no large differences shown, with the exception of the Czech Republic for which both surveys present the same curve, but this is translated forward one group in EU SILC.

263. For the remaining eleven countries (Denmark, Finland, Germany, Greece, Hungary, Latvia, Norway, Poland, Portugal, Spain and Sweden) different income group distributions are seen in ESS and EU SILC results. Some countries present discrepancies more precisely for one or two deciles, for example Germany, where the main difference between the two surveys can be seen in the first group (10% for ESS and 18% in EU SILC when the other percentages for upper deciles are similar). But for some other countries, the distribution is completely different from one survey to the other. For instance, the case for Portugal where ESS results show a 'bell curve' and EU SILC ones show a flat curve across the groups.

264. Again results are not encouraging, and the presentation of this ESS income group data as accurate population estimates without adjustments is not recommended, neither is the use of ESS data for the calculation of income poverty estimates (using amid-range method) however broad.

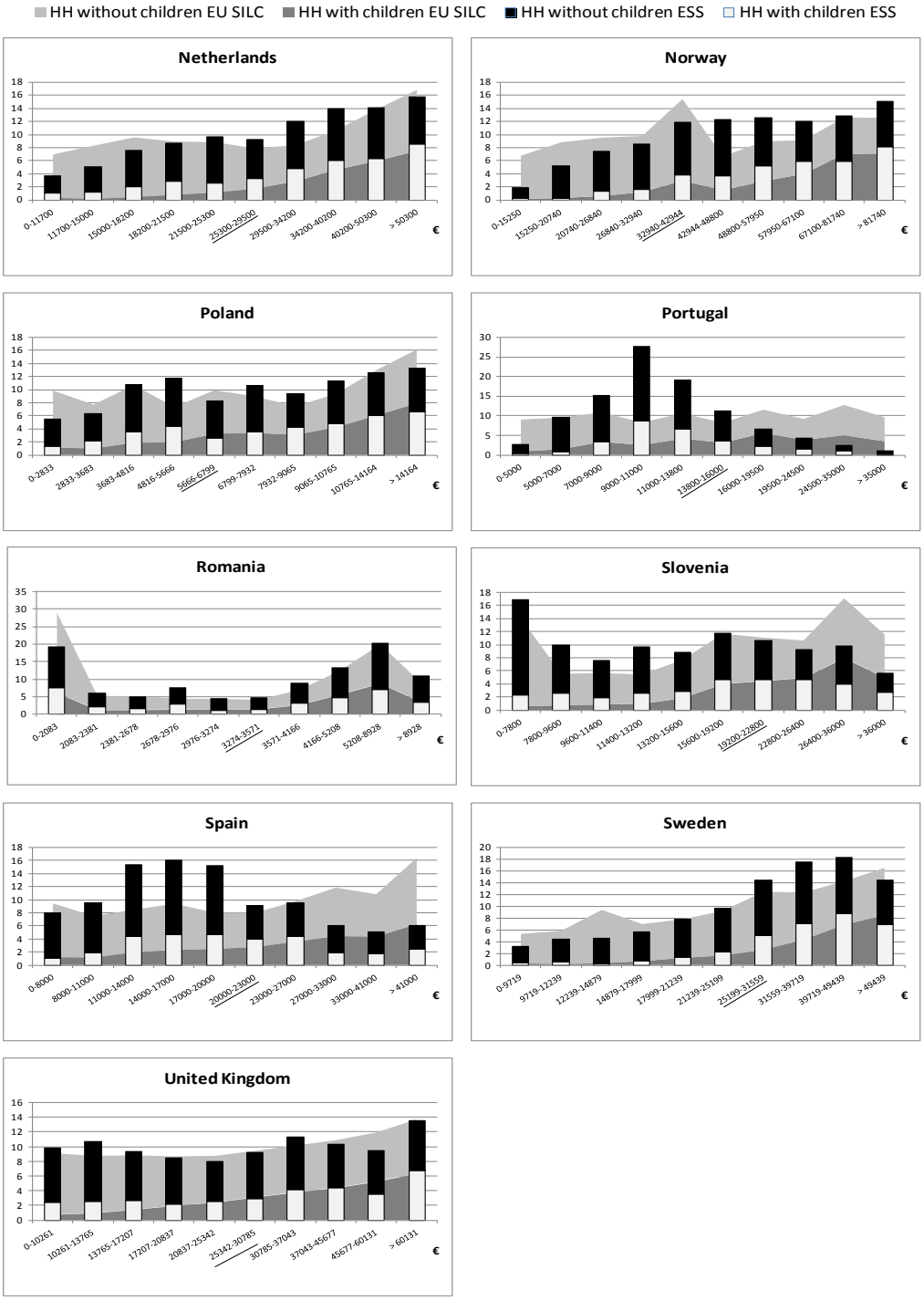
Figure 9. Household income distribution in ESS and EU SILC by household type, 2008.



Note: the median income is in the underlined group.

Source: Author's calculations of ESS and EU SILC, 2008.

Figure 9. Household income distribution in ESS and EU SILC by household type, 2008 (cont.)



Note: the median income is in the underlined group.

Source: Author's calculations of ESS and EU SILC, 2008.

3.5.4 Smoking, drinking and drug use rates in HBSC and ESPAD

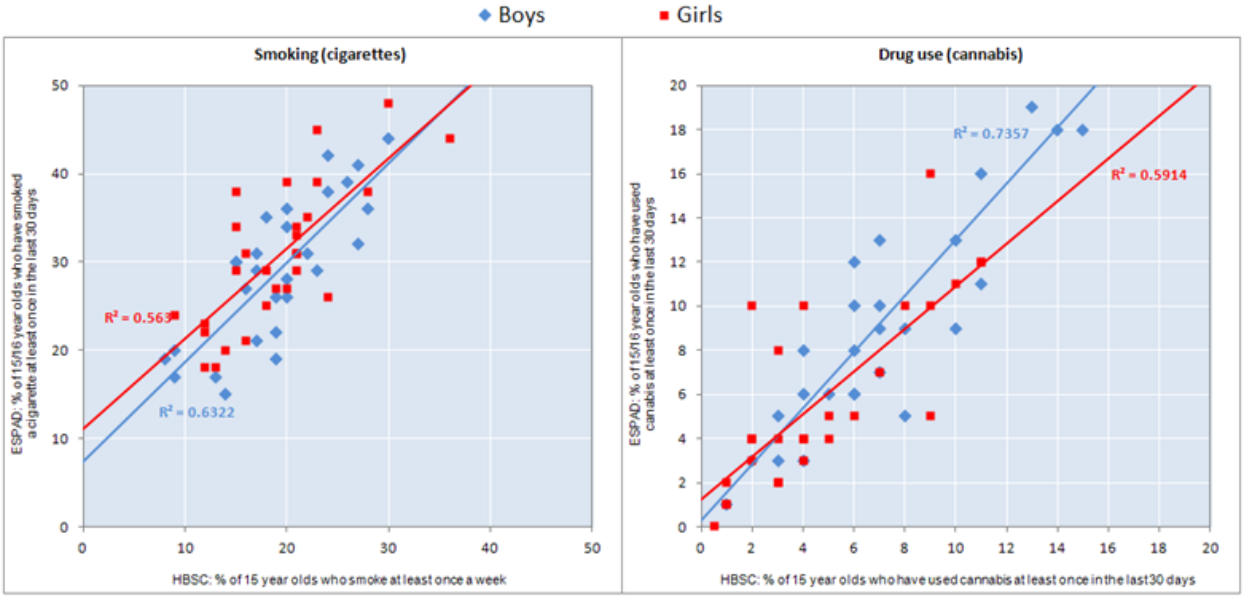
265. An important dimension of child well-being is risky behaviours, and especially substance abuse. Both the HBSC and ESPAD surveys cover a range of information on this subject, which can be used to compute comparable indicators on the prevalence of smoking cigarettes and cannabis use. These comparable indicators allow for a simple validity test between these two child surveys by measuring the consistency of outcomes related to these risky behaviours.

266. The correlations between the proportions of children that smoke cigarettes and use cannabis by gender, as reported by HBSC and ESPAD, are shown in Figure 10. Obvious differences in reporting methods between the two surveys (differences in the target population, and the frequency and time range of responses) create certain limitations. First, the HBSC survey covers children aged 15 years of age while ESPAD measures outcomes for children aged 15 or 16 years; and second, the time reference and prevalence in the smoking indicator varies (smoke at least once a week in HBSC, compared to in the last 30 days ESPAD).

267. Despite these limitations in comparability, the correlations provide an indication of the reproducibility of substance abuse estimates at the national level by gender, with results for boys being more consistent.

268. Overall, the prevalence of cannabis use is most closely matched with a correlation coefficient of r=0.86 (R^2=0.74) for boys and r=0.77 (R^2=0.59) for girls, which may reflect the similarity in the questions asked. However, the correlation for the prevalence of smoking is also high with a correlation coefficient of r=0.80 (R^2=0.63) for boys and r=0.75 (R^2=0.56) for girls. The high level of correlation between the indicators on smoking and drug use for both genders, despite the definitional differences, provides external reliability for both measures, and leaves researchers with a certain level of confidence about the ability of these respective items to capture underlying smoking risks in the child population.

Figure 10: Associations between smoking prevalence and cannabis use as reported in HBSC and ESPAD



Source: Authors calculations of the most recent waves of HBSC and ESPAD.

269. Although HBSC measures the proportion of children who have been drunk at least once in their lifetime, while ESPAD measure the proportion who have been drunk at least twice. The correlation between the surveys when measuring the prevalence of alcohol abuse is much weaker with a correlation coefficient of r=0.35 (R^2=0.12) for boys and r=0.26 (R^2=0.07) for girls (not presented in the figures). While to some extent this can be explained by the definitions difference the low level of correlation suggests these measures do not capture the same underlying risk factors.[24]

3.6 Summary discussion

270. There are three key points to be made on the basis of the analysis. Children's self-reported data are reliable between surveys and over time – and are robust to minor changes in method and question. Child surveys meet the standards that are implicitly accepted by many adult surveys in method and result. And, finally limits to these surveys need to be respected, including who doesn't join, and who, when in the survey, doesn't answer which questions.

271. In the past five years researchers have produced a wealth of child indicators covering a range of non-monetary topics designed to inform policy about the quality of children's lives. The evidence above suggests that amongst those indicators identified as key to children's well-being outcomes, almost all represent reliable estimates of the experience of the child population in European and OECD countries.

272. An important finding is that reliability in the response gleaned from children, between surveys, does not differ to any great degree from that shown for adults – in fact evidence suggests that child derived indicators between surveys may be more robust (this could be due to the similarities in good practise data collection between child surveys). Changes to the same surveys across the waves also do not seem to dramatically change results.

273. Nevertheless, improvements can be made in the calculation of indicators for the purpose of international research and monitoring of child well-being (see section 5.5). First, whilst appreciating many child well-being items are drawn from background questionnaires to surveys, is that non-response affects the accuracy of estimates and so comparison, and the adjustment to weights used to represent pupation estimates should be used to account for some of the bias resulting from non-response. Second, hidden populations, such as children not-enrolled in school, will affect the ability of the surveys to unbiased population estimates. Much work can be done to improve the capture of hidden populations in child surveys.

[24] In the 2005/6 and 2009/10 data researchers will be able to use the ESPAD indicators in HBSC, as the survey now have the same questions (most in the mandatory, and some in optional modules).

CHAPTER 4. GAPS IN CHILD COVERAGE AND INTERNATIONAL CHILD WELL-BEING DATA

A common criticism of recent international comparisons of child well-being have focussed on the missing aspects of well-being, and that selection of indicators into well-being frameworks have been driven by what's available, as opposed to what's preferable. Various critiques have identified child protection and neglect, neighbourhood factors, civic and political participation, and mental health as important topics yet to be included in comparisons. Moreover, commentators have highlighted the adolescent focus in the data presented in multidimensional frameworks, and the lack of information and application of child participation in these works. The purpose of this chapter is to take stock of available data to inform the development of existing and future surveys, and to provide contextual information necessary for a fuller understanding of the limitations of the present international comparisons of child well-being. The results show that children's participation in international surveys is completely missing before the age of nine, that there remain important gaps in the knowledge of children's well-being, by dimension, in particular in the coverage of at-risk children and the area of child protection and neglect. And that the cross-sectional nature of the survey data means little is known about the persistence of experiences (or development) per child in each of the dimensions.

4.1 Introduction

274. Identifying gaps in international child well-being data serves two purposes. First, by showing the gaps in the coverage of well-being indicators by concepts provides vital information for future development of surveys designed to inform cross-national monitoring of children. Second, identifying gaps in the data allows for a greater understanding of how to appropriately interpret the available data for monitoring the well-being of children. To provide robust information for policy-makers, it is important to be clear not only about which children the data refer to, but also what aspects of child well-being the data measure and, moreover, what is missing.

275. Gaps will be identified by mapping child well-being indicators available in the international surveys by broad dimensions based on socio-demographics, public policies for children (education, health, income, housing and environment, and civic participation and time use), children's relationships and subjective perceptions of life in different domains. Policy dimensions, relationships and subjective perceptions of life will then be cross-referenced to the socio-demographic characteristics of children (such as age and ethnicity). Together these data will provide a reference point for the cross-country indicators of child well-being by concept and child-coverage, and importantly where gaps in the knowledge lie.

276. Indicators categorised by child well-being dimensions will be introduced and summarised in terms of who provides the data (respondents) and the child age coverage. For each dimension there will also be a discussion of weaknesses in the data collection and missing concepts.

4.2 Methodology

277. To assess what indicators are available from cross-national surveys of children the first step was to review the background questionnaires[25] published by each survey and across each wave, and the second was to identify how these responses are collected and translated in to data points in the survey datasets.

[25] This review did not include non-compulsory questionnaires, such as the PISA parental questionnaire.

This data is then used to categorise the information into dimension and domains within dimensions. Each domain is then described, before a frequency analysis provides a summary of each dimension.

4.2.1 Indentifying dimensions and domains of child well-being

278. Dimensions have been defined after collating all questionnaire items into a database, and categorising these items into over-arching policy groups, socio-demographic items, relationship items and subjective perceptions of well-being. Items with dimensions have then been re-ordered into commonly recognised domains.

279. Given the objectives of the work the obvious choice was to separate out socio-demographic characteristics before assessing the remaining data by policy amenable fields and on the basis of subjective perceptions of quality of life. For the purposes of the chapter, subjective perceptions are considered those that provide opinion or a personal reflection on a situation or behaviour. Objective measures are those where respondents are asked to provide factual information regarding behaviours, circumstances or experiences. In this sense the framework broadly follows the suggestions outlined in OECD (2009) and the Stiglitz report (2010).

4.2.2 Analysis of available data

280. Using the frequency analysis conclusions can be drawn regarding the availability of data by dimensions and domain, and weak points in collection can be identified as well as conceptual gaps in the surveys.

281. The paper also reviews the items and their form, for the purposes of assessing the analytical capacity of the data collected. In the frequency analysis the number of scale, ordinal, nominal and binary indicators are compared for the future purposes of assessing the precision by which child well-being data is collected, and recommending appropriate forms of analysis for that data.

4.3 Socio-demographics

282. The first section reviews the socio-demographic data available from the surveys reviewed. It organises the discussion in to seven domains of age, family form, languages, migrant status, parental education, religion and sex.

4.3.1 Socio-demographic domains

283. The domains in the socio-demographic dimension are selected on the basis of common child characteristics (age, sex, etc.), family backgrounds (form and parental education), and non-income related family behaviours (language use, religion).

- In the case of the three households surveys (EQLS, ESS and EU-SILC) the **age of the children** in the household is collected by asking the most knowledgeable adult (head of household), and can be reported as years or years/months of age. In the child surveys, there is no variation in months or weeks identifiable in the household survey results, unless birth dates are provided. The child surveys are more precise, asking either for month and year of birth (TIMSS, PIRLS and HBSC), or for the birth date of the child (PISA).

- **Family form** in the all of the surveys includes data on the relationship status of parents, and whether both parents and brother and sisters live in the household. On occasion extended family members are included. The three household surveys include a review of all members of the households and marital status of the household head. The child surveys can also include the

members of the family that are step-parents (ESPAD and HBSC and PISA), or a simple count of the number of family members and the number of children (TIMSS and PIRLS).

- In the child education surveys (PIRLS, PISA, TIMSS), which require the completion of cognitive tests; the different **languages spoken in the home** are commonly recorded. For PISA a distinction is made between the test language and a list of other languages (national-specific selection). For TIMSS and PIRLS not only are there questions about speaking the test language, but also questions about the time spent speaking the language at home, and more specifically the time spent speaking the language with adults in the home (PIRLS). In HBSC, questions on languages used in the home are included in the optional package asked in 11 countries.

- Items recording the **migrant status** of the child or family are a regular addition to the surveys, with the exception of HBSC, who include this in the optional package undertaken by around one-third of countries in 2009/10. The question is asked in several different ways, the simplest of which is found in ESPAD, put as 'Where were you born?' The households surveys all contain questions about the respondents' place of birth and the respondents' parents' place of birth and so second generation migrant status information is available. PIRLS and TIMSS ask if the child and their parents were born in country of testing, splitting the parental question for mother and father separately (in contrast, PISA asks about parents in general - though a country specific list is provided). Additionally TIMSS, PIRLS and PISA ask children their age when they entered the country if they were born outside the country, which provides additional detail needed to analyse the time the child has had to integrate into the country of test. Between 2003 and 2007 there were minor changes to the way in which these questions were asked in the TIMSS survey.

- **Parental education** is included in all of the household surveys as well as TIMSS and PISA. ESS asks respondents for their education level, their partners and their parents, as does EU-SILC. With this data, intergenerational education transmission can be tested over two generations against the outcomes experienced by the children in the household. ESPAD asks pupils for their mothers and fathers education separately, as does TIMSS (using the ISCED scales). Items covering education levels in PISA have changed across the surveys, though the information is always collected by ISCED level.

- Questions of **religion** have only been included in one cross-national survey in the OECD and European region in recent years. The European Social Survey asks objective questions about membership of a religious group, which religious group, and follows this up with a subjective questions about 'how religious' the adult respondent considers themselves to be.

4.3.1.1 Socio demographic summary

284. There is very little left unasked in terms of the socio-demographics of children in the surveys. Overall, however, data on the children's language use at home and religious practises are least requested. Moreover socio-demographic data collected on all household members in household surveys mean that estimates of the age and sex of children, by population, is readily available to European researchers.

285. In terms of gaps, information is not always available on extended family or the ages of siblings in child surveys (for birth order analysis). Information on cultural (non-religious) practises may also be of interest.

Table 16: A summary of socio-demographic indicators

	Age	Family form	Languages	Migrant status	Parental education	Religion	Sex
Number of indicators	7	19	6	26	15	2	6
Number of sources	6	9	3	7	4	2	4
Adult data (parents)	4	10	0	8	7	1	1
Child data	3	9	6	18	8	1	5
Child ages	9/10; 10; 15	10; 11, 13 and 15; 15	9/10; 9-13; 15	9/10; 9-13; 10; 15	13; 15	15/16	9/10; 9-13; 10; 15
Migrant	7/7	17/19	6/6	26/26	15/15	2/2	6/6
Scales	5	11	2	3	1	0	0
Ordinal	1	0	0	4	9	0	0
Nominal	0	4	1	7	0	2	1
Binary	1	4	3	12	5	0	5

4.4 Policy dimensions

286. This section reviews the objective, non socio-demographic, items asked in each of the surveys and organises the data in policy dimensions. In this sense the framework broadly follows the suggestions outlined in work by Bradshaw and colleagues (2007), OECD (2009) and the Stiglitz report (2010). Each dimension contains information organised by domain before these results are summarised and missing concepts discussed.

4.4.1 Education and schooling

287. The dimension 'education and schooling' has been organised into the domains of: general education, attainment and achievement, attendance, early years attendance, extra curricula activities, homework, and teaching and learning.

288. Selection of domains in the education and schooling section were made on the basis of the specificity in the items to main characteristics of the school experience, from broad education questions, to achievement and attainment, engagement with the systems (attendance), experiences when engaging with the system (teaching and learning), and learning outside of the system (home work and extra curricula).

- **General education** items are included mainly in the household surveys and used to identify the education levels of the head of households. EQLS and ESS include items about the length of time spent in education, age of completion, the highest level of attainment, and the field of study. EU-SILC asks about the education levels of all members of the households. In 2006, PISA also asked children about the different types of science lessons they attended (compulsory and non-compulsory).

- **Attainment and achievement** levels of children are collected by PISA, TIMSS, and PIRLS via the cognitive testing in their surveys designed primarily for international comparison of educational achievement. These surveys also ask about recent school marks and the grade the child is attending (PISA samples children by age, TIMSS and PIRLS sample by grades). ESPAD asks a similar question about the educational grade of the respondent, as well the quality of their school work.

- **Attendance and 'early years' attendance** are important questions to test the child's engagement with school and to identify, where possible, the effects of early years interventions on children's life outcomes. PIRLS includes a simple question asking if children attend school. ESPAD and

PISA ask instead about the number of lessons missed in the recent weeks or week. Expanding the attendance information PISA also asks children about the number of students in their class – to represent broader attendance (PISA sampling does not cluster class groups and so numbers are not easily estimated).

- PISA is the only child survey that asks about **pre-school attendance** (see general education above for the household survey approach) as well as the age when started primary school. Annex 6 shows that the vast majority of children in the EU and OECD regions are confident with answering the question about attendance of preschool.

- **Extra curricula activities** are included in the PISA background questionnaire and ask students about their attendance of extension course, remedial course and private tutoring. TIMSS also asks children if they have taken extra lessons in mathematics or science (although the level of the teaching is not specified).

- Children's experiences of **homework** have been positively and negatively associated with educational outcomes. Homework is also linked to time use and school pressures experienced by children. PIRLS asks children about the amount of reading homework they get, how often they get it, and who helps with it. PISA in 2000, asked how children's homework was contributing to class work (counted as part of marks) their interest in homework, where they did it, how much they did, and with what distractions (for instance, if the TV was on). The questions about homework in TIMSS simply ask how often children get mathematics and science homework, and the length of time it takes to complete it (no reference to the amount of homework).

- The final education domain is **teaching and learning**. This domain captures interaction with teachers and classroom environments, including teaching practises. ICCS asks students about whether the teachers encourage the students and their peers, whether other students speak in class about political issues, open discussion and opposing opinions, and whether teachers represent all sides of an argument as part of teaching. PIRLS asks two questions about teaching practises, including the various people who might read aloud in class, and the follow-up work associated with the reading (for instance writing a summary of the piece). The PIRLS questions were reduced and simplified between 2001 and 2006. Finally, PISA asks about the classroom environment in terms of peer behaviour and interaction with teachers, as well as student's use of the school facilities (library, computers, and laboratories).

4.4.1.1 Summary of education and schooling items

289. The most common data in the education dimension are data for general education and teaching and learning practises. The former is asked almost exclusively to adults (and so applicable for indicators of home or parental educational environment or intergenerational transmission of education) and the latter is almost exclusively asked of children (see Table 17). Early years experiences are least covered, and only in PISA, which, given the recent focus on early years experiences for later child outcomes, is underutilised. Age coverage in the data is relatively good given that the educational surveys cover the youngest to the oldest child (9 through 15), but there remains a large gap in the knowledge about educational experiences in the middle and early years of childhood.

Table 17: A summary of education and schooling indicators

	General education	Attainment & achievement	Attendance	Early years	Homework	Extra curricula	Teaching & learning
Number of indicators	7	9	5	2	7	3	12
Number of sources	3	3	4	1	3	2	3
Adult data (parents)	6	0	0	0	0	0	0
Child data	1	9	5	2	7	2	12
Child ages	15	9/10; 15; 15/16	9/10; 9&13 15; 15/16	15	9/10; 10; 15	10; 15	9/10; 13/14; 15
Migrant	7/7	9/9	5/5	2/2	7/7	3/3	6/12
Scales	2	4	3	0	4	1	10
Ordinal	2	4	0	0	1	1	2
Nominal	1	0	0	0	0	0	0
Binary	2	1	2	2	2	1	0

4.4.2 Health and risks

290. The domain 'health and risks' has been organised into the domains of: general health, body and image, mental health, disability and chronic illness, nutrition, personal care and injury, general risk behaviours, and the risk behaviours of drinking, drugs, sex and smoking.

291. Selection of domains in the health and risks dimensions follow commonly understood groupings of health outcomes and health behaviours. Outcomes are covered in terms of illnesses and morbidity (from general health, mental health and chronic illness and body image [BMI]), and behaviours include those that are positive (such as nutrition and physical activity) and those that impose risks.

- **General health information** asks children, or the household respondent, for an objective assessment of their general health (different from subjective assessment of feeling). EQLS and EU-SILC ask about the state of the person's health or general health, as well as asking if health issues have limited the person's activities (EU-SILC). The household survey information is useful for identifying adult experiences that may affect the home life of children including in terms of: caring responsibilities or longer term experiences of unemployment and poverty (if adult is work restricted or on incapacity benefits). From the childhood surveys, HBSC collects information on the prevalence of certain health symptoms, asking children whether they experience: headaches, stomach-aches, back aches, feel low, irritable or have a bad temper, feel nervous, feel dizzy or having difficulties getting to sleep (from almost every day to rarely or never). Some of these HBSC items are useful for indicating mental health problems and the set is scalable into somatic and psycho-somatic composites.

- More detailed items are also available on **mental health and chronic illnesses**.[26] The household surveys all ask questions on the subject. ESS and EU SILC ask about long standing illness disability (chronic illness), and EQLS includes an item on chronic mental health problems (although the question is not separated out by health problems). ESPAD asks children about mental health symptoms, such as losing their appetite, feeling depressed, pressured or sad, and not be able to work (school work) or having difficulty concentrating. A second ESPAD question directly asks children how often they have self-harmed or attempted suicide.

[26] Optional packages in HBSC cover physical disability and chronic conditions (14 countries in 2009/10) parts of which are being reviewed for inclusion in the mandatory questionnaire in 2013/14. Other optional packages include medicine use (18 countries in 2009/10), Global well-being/KIDSCREEN (18 countries in 2009/10), and the Strengths and Difficulties Questionnaire (5 countries in 2009/10).

- Information on **size and weight and body image** is collected in the HBSC survey and used for Body Mass Index (BMI) calculations, and subjective perceptions of weight, and dietary behaviours. The information on BMI is asked of children through two questions about their weight (reported in pounds [lbs] or kilos [kg]) and their height (in metric or imperial scales). The data on weight and height are notoriously unreliable, and evidence from previous HBSC surveys suggested there is clear bias in the non-reporting of this information country-to-country (Currie *et al*, 2004). Another HBSC question in this domain asks children whether they are dieting or trying to lose weight; responses to this question range from a simple 'yes' to three different 'no' options (I'm fine, I'm too thin, no – but I should).

- Both **nutrition** and **personal care and injury** are objective health indicators linked to mental and physical health and body image. The only source for information of this type is the HBSC survey. In terms of nutrition, three separate questions address the regularity of taking the main meals of the day (breakfast, lunch, and supper, although only breakfast is included in the most recent mandatory questionnaire). A further set of nutrition questions include: the regularity of consuming sweets, soft drinks, alcoholic drinks, fruits, and vegetables. The personal care data collected by HBSC looks at dental care (how often children brush their teeth, from more than once a day to never), physical activity of over 60 minutes per day (in the most recent week), and whether a child reports have received medical attention for an injury in the past year (by number of times to four or more). All questions are asked of the three age cohorts in the survey, aged 11, 13 and 15.

- Two surveys cover in detail the risk behaviour undertaken by children and adolescents; these are ESPAD and HBSC. **Risk behaviour** information is covered in the sub domains of drinking, drug taking, sex and smoking. General questions of risk behaviour tend to cover all of the above mentioned aspects of risk and address personal experiences and experiences of peers and siblings.

- For indicators of **drinking,** items cover access to alcohol, age of first use, regularity of drinking alcohol, and experiences of drunkenness, binge drinking in the past month (consumption of more than 5 drinks in a row), and experiences of what happens to youth after drinking (including getting in trouble with police, feeling happy or depressed, changes in behaviour that lead to further risks or injuries). ESPAD also asks where the child was when they last drank alcohol.

- ESPAD also explores, in detail, youth experiences of **drug taking**. For cannabis, issues of access, regularity of use, age of first use, and subsequent behaviour/trouble are all addressed. Added to the list is a peer-pressure question, which asks if children have ever refused to take cannabis. Beyond the use of cannabis, ESPAD also asks about experiences of using a range of other drugs including: sedatives, amphetamines, LSD, crack, cocaine, heroin, mushrooms, GHB and steroids ('relevin' is included as a dummy drug to test children's honesty in answering these questions). HBSC also asks about experiences of cannabis use (but this question is only asked of the 15 year old cohort).

- Fewer questions are asked about **smoking** in the surveys. Although in ESPAD questions about access and regularity of use are included. A parental relationship question is also included by ESPAD, asking children if their parents would allow them to smoke. HBSC asks two questions on whether the child has ever smoked, and the regularity of smoking at the moment. Neither survey asked the children how old they were when they smoked their first cigarette until HBSC included this as mandatory (along with drinking and drunkenness) in 2009/10.

- Finally, **sex** as a risk behaviour is included in the HBSC survey. The questions are only addressed to 15 year olds, and ask if they have had sexual intercourse, how old they were the first time, and the protection used by themselves or their partner during their last experience.

4.4.2.1 Summary of health and risks items

292. Child well-being health indicators are dominated by risk behaviours. ESPAD and HBSC both have a number of these measures, and particularly ESPAD focuses on this type of measure. Risk behaviours are generally asked of children in the mid teenage years, but results from HBSC shows that smoking and drinking can begin in middle childhood. Mental health question remain few and far between for children – more questions are directed to adults – although these and may be produced in educational surveys under the headings of stress or pressure in school (from teachers and/or parents).

293. In terms of the age coverage, because of the absence of health or risk behaviours in the PIRLS data, the age range for children covers 11 to 16 (see Table 18). Early years' health information and middle years' health information cannot be compared across countries. Also, because of the dominance of HBSC in this area, there is a limit to how far this data can be disaggregated to reflect the experience of children born outside of the host country.

294. Examples of broader conceptual gaps include adult health behaviours, medical visits, and unattended injuries. None of the surveys covered collect data on the risk behaviours of adults, which can impose pressures on the child home environment, income and safety (and as such could reflect child protection issues, as might unattended injuries). Moreover, beyond medically attended injuries, there are no other data available on the regularity of seeing health professionals – information of visiting the doctor for illnesses and check-ups would provide for a better understanding of how children from different backgrounds access health care (and in turn how this relates to broader life outcomes).

Table 18: A summary of health and risks indicators

	General health	Mental health & illness	Body & image	Nutrition	Personal care & injury	Risk behaviour: general	Risk behaviour: drinking	Risk behaviour: drugs	Risk behaviour: sex	Risk behaviour: smoking
Number of indicators	4	5	3	4	3	4	10	12	4	7
Number of sources	3	4	1	1	1	2	2	2	1	2
Adult data (parents)	3	3	0	0	0	0	0	0	0	0
Child data	1	2	3	4	3	4	10	12	4	7
Child ages	11, 13 and 15	15/16	11, 13 and 15	11, 13 and 15	11, 13 and 15	11, 13 and 15; 15/16	11, 13 and 15; 15/16	15; 15/16	15	11, 13 and 15; 15/16
Migrant	1/4	5/5	0/3	0/4	0/3	3/4	8/10	11/12	0/4	5/7
Scales	1	0	3	4	3	1	2	3	1	3
Ordinal	2	2	0	0	0	2	7	8	0	3
Nominal	0	0	0	0	0	0	0	0	0	0
Binary	1	3	0	0	0	1	1	1	3	1

4.4.3 Income and deprivation

295. The domain 'income and deprivation' has been organised into the domains of: general income and benefit income, households support, taxes, household costs, household debts, general deprivation, deprivation in terms of culture, education and food, employment security, and parental employment and conditions.

296. The dimensions of income and deprivation are selected into three broad categories reflecting of income and costs (and so includes debts and taxes for instance), various forms of deprivation, and issues related to employment and employment security (and so the main underlying source of income in the majority of cases).

- Information on **general income and benefit income** is critical for the assessment of income poverty in households with children. This information is considered most reliable when collected from adults, and as such is mainly collected in the household surveys. EQLS, ESS and EU-SILC all ask questions about the households' disposable income in net terms; and EU SILC asks for the same information but before taxes and transfers. ESS is also interested in the sources of income in the household, and provides a simple tick box for respondents to identify the main source of income in the home (using the breakdown of income sources in EU-SILC it is possible to derive the same information from this survey). EU-SILC also asks information about the income received from children under the age of 16, as well as for sources of benefit income (for instance education related allowances and disability payments). ESPAD is the only child survey to include an item directly related to money, which asks for the amount spent on the child's personal needs, with parental control, per week.

- Related to income are household supports received from, or provided to, relatives. These 'non-official' sources of income can be of varying importance country to country (for instance in Turkey around 20% of households benefit from this type of support [Rose and Özcan, 2007]). EQLS asks respondents if they give or receive regular support to or from relatives.

- Each household's **tax burden** is considered in the EU-SILC datasets. Questions include information on tax adjustments, tax on incomes and social contributions, and taxes on wealth. These data are important for assessing the net effect of tax systems on the poverty rate experienced in households with children.

- Household **costs and debts** are covered by EU-SILC and can be used to assess broader financial pressures on the households in which children live. The household's main respondent is asked about burdens in terms of total housing costs and interest payments (in scale terms), as well as arrears on loans, mortgages and utility bills (simple yes/no responses). EU-SILC also asks respondents for an assessment of the minimum monthly income required to make ends meet in the household. EQLS too, asks about household debt, again in terms of rent or mortgages and utility bills, but a time limit is given (experiences in the past 12 months).

- **General deprivation** in the surveys reflects on whether households can make ends meet on their present incomes in the ESS (an objective assessment of going without, rather than income required as above); and in EU SILC it is covered by items on the ownership of: a car, a colour TV, a telephone, a washing machine, as well as being able to afford to take a week's holiday per year. In EQLS, general deprivation is addressed by asking for yes/no answers to a list of affordable items (if wanted) including: keeping the house warm (see housing conditions below), replacing damaged furniture, taking a week's holiday, having protein meals, having friends over once a month, and buying new clothes. HBSC, PIRLS, PISA and TIMSS also ask children about general deprivation at home. The education surveys tend to include a check list of items that mix leisure items (i.e. computers games) and educational items (calculators etc.) with cultural items (i.e. musical instruments or books). Both PISA and TIMSS include questions about ownership of a vehicle in the home (as does HBSC). HBSC is the only survey to ask children if, and how many, holidays (vacations) have been taken with their family in the past 12 months.

- Other forms of deprivation; including cultural, educational and food deprivation, can be supplemented with many of the items included in the general deprivation questions in the child surveys. **Cultural items** refer to the number of books in the home (works of art or musical instruments are covered in the general lists). These questions are asked only in the educational surveys, and response options are grouped into numbers of 1 to 10 through to 200 or more in PIRLS, and 500 plus in PISA. Children are helped with answering this question by a note to count 25 books per shelf of books in the home in TIMSS, and 40 books per meter in PISA. There are very few missing responses on this measure for children aged 15 (see Annex 6).

- Items specific to **educational deprivation** refer to ownership of a computer (or in HBSC numbers of computers in the home), or ownership of items for learning, including a space to study, dictionary and calculator (in TIMSS). Three surveys ask about **deprivation of food** at home unique questions. EQLS asks if the household has run out of money for food in the past 12 months, EU-SILC asks if families can afford a protein based meal, and HBSC asks how often go to school or go to bed hungry.

- **Parental employment** is of particular importance to understanding the risk to child well-being experienced across countries. Unemployment can bring with it poverty risks and deprivation that predisposes a child to poorer present and future outcomes. Parental employment can also bring strains on the household and home environment, and crowd-out caring opportunities and leisure opportunities. Each of the household surveys asks about the respondents' employment and employment of at least one other member of the household (usually the partner). EQLS asks if respondents have ever worked, and EU SILC, EQLS and ESS ask about the number of hours worked by adults in the household. HBSC and PISA also ask children directly about parental employment. In HBSC the questions address whether adults living in the home work, where they work, and what they do (the latter two are both open-ended questions); and finally if they don't work, 'why not?' PISA employment questions have changed between the waves: in 2000, children were asked each parents main occupation and what they were doing at the time of interview, in 2006 this had changed to questions about mothers and fathers employment status.

- **Employment security** questions add to the knowledge of household conditions, because of strains on working parents, and their ability to plan for the future. Employment security in terms of contract type is covered in EQLS and ESS. ESS extends the questions to cover length of time seeking work if unemployed, the last year employed, and expectations for employment and under-employment if finding a job in the coming years.

4.4.3.1 Summary of income and deprivation items

297. The coverage of indicators by income, deprivation and by employment is quite evenly distributed. And partly to do with conceptions about the reliability of children's responses to income questions, the deprivation and employment indicators are more evenly split between adult and child respondent than income questions. Indeed, much of the underlying information which is used to calculate child poverty estimates, and provide information which complements estimates of household disposable income (how much of that is paid out in debts for instance), are collected at the household level. Again the youngest children covered by surveys (those aged 9 in PIRLS) are represented only in the deprivation figures (parental employment data is not collected by PIRLS).

298. The broader conceptual gaps in the income and deprivation dimension could include deprivation of items for leisure or sport (a bicycle for instance), for socialising (clothes), or for a broader engagement with school (uniforms, bags, or missing school events or outings). Pets in the home are increasingly viewed as important for children, and this item is missing from all international surveys of children.

Table 19: A summary of income and deprivation indicators

	Income & benefits	Support	Taxes	Costs	Debts	Deprivation (general)	Deprivation (cultural)	Deprivation (education)	Deprivation (food)	Parental employ	Employment security
Number of indicators	14	2	3	3	4	14	3	3	3	23	4
Number of sources	4	1	1	2	2	7	3	3	3	5	1
Adult data (parents)	13	2	3	3	4	8	0	1	0	11	4
Child data	1	0	0	0	0	6	3	2	3	12	0
Child ages	15/16	-	-	-	-	9/10; 11,13 and 15; 15	9/10; 15	9/10; 15	9/10; 15	11,13 and 15; 15	-
Migrant	14/14	2/2	3/3	3/3	4/4	12/14	3/3	2/3	2/3	19/23	4/4
Scales	10	0	3	1	0	2	1	0	0	3	1
Ordinal	0	0	0	2	0	3	1	1	1	0	2
Nominal	3	0	0	0	0	0	0	0	0	8	0
Binary	1	2	0	0	4	9	1	2	2	2	1

4.4.4 Housing and environment

299. The dimension 'housing and environment' has been organised into the domains of: main home, housing conditions, housing security, tenure, locality, crime, environment.

300. Housing and environment domains are selected to separate out immediate environments and conditions (so the home, the HBSC main home question is included here as it identifies children with multiple homes), and local environments and conditions (areas in which live children live and their conditions).

- **Main home** information is included in HBSC and asks children about the time spent in their main home and second home (if parents are living apart). The survey also asks children if they live in social care.

- **Housing conditions** are covered in three surveys, EQLS, EU-SILC and HBSC. EQLS asks about the number of rooms available in the household excluding kitchens, bathrooms, halls and storerooms, and rooms used for business. EU-SILC asks about whether the household is kept adequately warm, whether there is damage to the house, and whether the household includes a bath or shower. HBSC asks children if they have their own bedroom (which can be linked to the deprivation items asked in the other child surveys). **Housing security** questions are also in the EQLS survey, the question addresses the likelihood of the respondent having to leave the property in the next 6 months because of not being able to afford to live there. Related to security is **housing tenure**. EQLS and EU-SILC ask respondents if they are buying their home, or whether they own or rent it (social housing or privately), and in the case of EU-SILC, the dwelling type also.

- Broader questions about the access to certain services in the **locality** and the **size of the locality** are asked in EQLS and ESS. EQLS asks respondents if in their area there is: a food shop, a recycling facility, public transport, a post office, a bank, and a cinema. Both surveys ask about the description of the locality, ranging from a rural or countryside area to big city or city suburb.

- **Crime** questions refer to the quality of the locality in which families live, as do question about pollution. ESS asks respondents about whether they have been a victim of burglary in the past 5 years, and EU-SILC ask if respondents have experiences of violent crime in their area. Both ESS and HBSC ask respondents whether or not they feel safe in the area in which they live.

- Questions about **pollution** are asked in EQLS and EU-SILC. The former lists possible complaints about the local area in terms of noise, air pollution, water quality and litter. There is some crossover here with crime, as this is included in the list of possible complaints, as is space for recreational activities (green spaces type question). EU-SILC asks if respondents experience noise from neighbours, pollution or environmental problems, and again, crime.

4.4.4.1 Summary of housing and environment items

301. The indicators are quite evenly spread between home and locality environments and conditions in this dimension. The majority of data come from household surveys, and are therefore asked of adults, which are perhaps more appropriate and given the home focus of the data. Nonetheless this leaves scope for a greater focus in child surveys for children's reflections on housing, environments and conditions. In terms of age-related data, because HBSC is the only child survey that provides specific data on the subject collected form children, the range covers only those aged 11, 13 and 15. As with all other data collections age specific experiences (which cannot be gleaned from age disaggregation in the household surveys) are limited.

302. The broader conceptual gaps in this dimension include the environment specific to children in the home, such as their room or private space, and the siblings they share this space with. Questions about the local environment from the perspective of children could be strengthened including from the perspective of children's facilities in the area (beyond green space or play facilities, such as youth clubs), and interaction with other children in the locality whilst out of school.

Table 20: A summary of housing and environment indicators

	Main home	Housing conditions	Housing security	Tenure	Locality	Crime	Environment
Number of indicators	1	3	1	4	3	4	3
Number of sources	1	2	1	2	2	3	2
Adult data (parents)	0	3	1	4	3	3	3
Child data	1	0	0	0	0	1	0
Child ages	11,13 and 15	-	-	-	-	11,13 and 15	-
Migrant	0/1	3/3	1/1	4/4	3/3	3/4	3/3
Scales	1	1	1	0	0	1	1
Ordinal	0	0	0	2	0	0	0
Nominal	0	0	0	1	2	0	0
Binary	0	2	0	1	1	3	2

4.4.5 Civic participation and time use

303. The dimension 'civic participation and time use' has been organised into the domains of: civic participation (political), civic participation (social), personal time use, education time use, friendships (time use), and religion time use.

304. Selection of domains in the civic participation and time use dimension are based on the division of civic and social participation in previous cross-national comparisons (see Bradshaw et al, 2007) and on the activities or environments in which children are likely to spend the majority of their time (home, school, or interacting with friends, family or the communities in which they live).

- Question on **political participation** are found in both household and child surveys. EQLS and ESS ask respondents about the whether they vote. ESS expands the question to include whether respondents contact politicians, signing petitions, how interested people are in politics, whether they trust public institutions, and whether they are a member of a political party. The ICCS

survey asks children about their national and international political interests, their plans to involve themselves in campaigning as an adult, plans to join a political party, a trade union or plans to stand in elections when they are old enough.

- **Social participation** questions are asked in CIVED, ICCS directly to children, and in EQLS and ESS to household respondents. The ICCS question is simply put 'How interested are you in social issues?' CIVED questions were more elaborate and included participation in non-governmental organizations for youth, human rights or the environment, as well as school groups and clubs (a supplementary question was also asked about attending meetings of these groups). ESS asks people about their voluntary activities in the past month, and the amount of social activities taken part in compared to people of the same age.

- Broader issues of **personal time use** are included in many of the surveys. The educational surveys focus on computer use, watching television, or undertaking reading for a hobby. Generally these questions include (as in the case for TIMSS) a broader set of activities such as listening to music, playing sport, or spending time with friends. ESPAD also asks children how much time they spend helping in the home (shopping, caring, cooking etc.). ESS asks the households respondent how many hours a week they spend watching TV, listening to the radio, reading newspapers and surfing the internet. HBSC include items on time use watching television and games playing.

- Education time use and friendships are closely related to personal time use data (which can often include relevant information as part of longer lists of activities). Questions specifically directed to **educational time use** are included in the PISA, PIRLS and HBSC surveys. HBSC asked about the amount of time spent on homework daily (split for weekdays and weekends – this has been dropped int eh most recent wave of 2009/10). PIRLS asks about reading homework and reading *for* homework, as well as time spent reading and being read to (various media are listed: magazines, books etc). PISA asks about time spent with science media (documentaries, websites, and books), the amount of time spent on homework in the three literacy areas, and the regularity of help with homework from relatives. CIVED and HBSC have asked children about the **time they spend with friends**. The CIVED survey asked about time just hanging out with friends, and time spent in the evening outside of the home with friends. HBSC asks similar questions about the time spent directly after school and in the evenings with friends. An additional question asked by HBSC collects data on the extent of interaction with friends through phone, text, or email messages.

- **Religious time use** is addressed in the EQLS, were adult respondents are asked how often they attend religious services (excluding weddings and funerals and other main religious events).

4.4.5.1 Summary of civic participation and time use items

305. Personal time use is by far the most requested information in this dimension. Children of all ages of the surveys contribute to the results. In contrast religious time use is least requested, and only asked of adults. Civic and social participation data will soon be available from the ICCS 2009, but this is only for children aged 13 or 14. Only HBSC can provide information of the amount of time children spend with friends for recent cross-national comparisons, this does mean however that disaggregation by migrant and non-migrant children is not possible (with the exception of countries completing optional packages).

306. Given that the surveys mainly cover older children it is surprising that no information is asked about children in paid employment. Broader socialising with family members could be reintroduced in the PISA surveys (or elsewhere) following the discontinuation of items on parental interaction since PISA

2000. Moreover there is no data on how children spend time with adults outside of the home (except inferred by religious practises). Surveys also do not ask how much time children spending eating or sleeping, which can impact on a range of health issues (physical activity however is included in health behaviours).

Table 21: A summary of civic participation and time use indicators

	Civic participation (political)	Civic participation (social)	Education time use	Friendships	Personal time use	Religion time use
Number of indicators	12	6	8	5	24	1
Number of sources	3	3	3	2	7	1
Adult data (parents)	6	3	0	0	2	1
Child data	6	3	8	5	22	0
Child ages	13/14	13/14	9/10; 11,13 and 15; 15	11,13 and 15	9/10; 9-13; 11, 13 and 15; 15; 15/16	-
Migrant	6/12	3/6	7/8	0/5	21/24	1/1
Scales	8	2	6	5	15	0
Ordinal	0	2	1	0	3	1
Nominal	0	1	1	0	3	0
Binary	4	1	0	0	3	0

4.5 Indicators of personal relationships

307. Personal relationships are not commonly conceived as policy amenable, and as such are treated separately from the policy dimensions. In the conceptual frame of well-being, the relationship indicators can be treated as contextual variables. The relationships dimension is organised into the domains of: family and social relationships, school relationships, and antisocial behaviour.

308. Selection of domains in this dimension broadly follows the time use approach, and so is based on the main environments in which children spend time and interact with others: at home (family), at school and when with their friends.

- In the household surveys questions about **family and social support** are asked to the head of households and so can only reflect on child well-being from the perception of broader social environments, or social capital indicators. For example, ESS asks household respondents if there is anyone with whom they can discuss personal matters; the response option is a simple yes or no. EQLS includes children in the list of contacts for support, written or phone contact (including emails) and general face to face contact (other people include: friends, neighbours, partners and siblings). Of the child surveys, only HBSC and ESPAD ask about positive interaction with family and friends (school environment tends to focus on the negative aspects of peer interaction – see below). In HBSC a question is included about how easy is it to talk to a range of people including parents and siblings. HBSC also asks about the number of friends a child has. ESPAD, in its most recent survey, asked children about how often they emotional support from a friend, and the frequency of certain types of interaction with their parents (including parents setting rules, knowing where the child is in the evening, lending money, receiving money as gifts, and getting emotional support from parents). PISA's contribution to familial relationships was undertaken in 2000 and has not been repeated since. Questions were asked to children about how often their parents talked to them about: politics, culture and media, about school, and about things in general; or whether parents listened to classical music with them, and how often parents ate their main meal at home with the children.

- The **school relationship** domain covers objective reporting of incidents in school to do with bullying and anti-social behaviour. HBSC ask students how often in the past couple of months they have been a victim of bullying, and how often in the past couple of months they have instigated bullying. In both TIMSS and PIRLS the same question are asked of children about the school experiences of theft, bullying, and fighting (occurring to both themselves and classmates). In each of these surveys questions of negative school relationships are regular items.

- Questions of **antisocial behaviour** are covered in the ESPAD and HBSC surveys. The questions are therefore asked of children aged 11 to 16. This section is separate from school relationships because questions reflect broader life context rather than experiences of bullying or fighting, teasing or stealing in school. HBSC ask students how many times they have been involved in physical fight in the past 12 months. ESPAD ask how often have children bullied, hit someone, stolen from someone, or damaged property. This ESPAD question is complemented by asking how often in the past 12 months the child has hit their teacher, been in a fight, used weapons, stolen something, committed arson, damaged property or got into trouble with the police.

4.5.1 Relationships summary

309. Relationships with family members are the most commonly reported relationship indicator in the surveys, with respondent type being evenly split between adults and children. And although evidence is available on the interaction children have in school and with peers, there is very little in the way of positive interactions in the school environment. Younger children are only asked about relationships in the school (PIRLS).

310. There are notable gaps in the relationships data given that children do not only interact with family, other children and teachers. Children are not asked about interactions with professionals other than teachers, or a broader set of adults outside of the home. For older children it is also appropriate to ask about interaction with employers (formally in a work setting, or when undertaking jobs such as babysitting), or with representatives in the community. For younger children, information on relationship with after-school care providers, older siblings and extended family is needed.

Table 22: A summary of relationship indicators

	Family & social relationships	School relationships	Antisocial behaviour
Number of indicators	9	4	4
Number of sources	5	3	2
Adult data (parents)	4	0	0
Child data	5	4	4
Child ages	11,13 and 15; 15; 15/16	9/10; 9-13; 11,13 and 15	11,13 and 15; 15/16
Migrant	7/9	2/4	3/4
Scales	3	2	1
Ordinal	4	0	3
Nominal	0	0	0
Binary	2	2	0

4.6 Indicators of subjective well-being and opinions

311. Indicators of subjective well-being are presented separately as personal reflections on levels well-being. This is commonly separated from objective perceptions of well-being (see OECD, 2009; of the Stiglitz Report 2010). This dimension is organised into the domains of: future aspirations, body image, education, health, life satisfaction, personal relationships, political opinion, risk behaviours, satisfaction with locality, school life, and subjective material situation.

312. The selection of the domains is based on subjective perceptions of indicators that, objectively collected, would be placed in the dimensions and domains outlined above.

- **Aspirations** refer to the child's perception of their future. There are four questions in total, asked of children aged 13, 14 and 15 in TIMSS, CIVED and PISA respectively. Each of the surveys asks children about their educational aspirations, though in different ways. CivEd asks about expected years of education, TIMSS about expected ISCED level at completion, and PISA, about expected final grade. PISA also asks a question about what job the child expects to be doing when they are 30 years of age.

- There are two questions on **body image** asked by the HBSC survey. One asks children about personal reflection on how good looking they are, the second is about body weight.

- By far the most common subjective opinions asked of children are about **education and school**. In total 18 questions are asked in the four child surveys of HBSC, TIMSS, PIRLS and PISA. Each survey asks for the child's opinion of learning or reading (depending on the surveys' subject of interest), complemented in TIMSS by questions on subjective perceptions of abilities in maths and science and how much children like school (a question also asked in the HBSC survey). PISA asks a specific question about how teaching of science in school might prepare the child for a science career, as well as the child's general interest in a range of science based subjects. Additional questions are asked in the HBSC survey to cover children's perception of school pressure, classroom environment (classmate support), and how teachers rate their abilities compared to their peers.

- Only two questions of subjective **health status** are asked in the child surveys. HBSC asks children to report their subjective health on a scale of excellent, good, fair or poor. ESPAD also ask the question as part of a broader set of life satisfaction questions (see below).

- **Life satisfaction** is an important child well-being indicator because it represents the overarching summation of the quality of life experienced by the child. Across the surveys this question is asked three times. In the household survey ESS the question is delivered to parents (or children over 15 if they are the household respondent). In the ESPAD survey, life satisfaction is asked as part of a list of questions including satisfaction with finance and health. Perhaps the most commonly used life satisfaction scale in international comparisons of children is that derived from the HBSC, this measures uses the Cantrill's ladder technique, which asks children to place themselves on a rung of the ladder based upon the opinions of general quality of life at the moment of answering the survey.

- **Personal relationships** are covers in the ESS and ESPAD. The questions ask respondents about how happy children are with their parents (ESPAD) and whether household respondents to ESS trust others, think people might take advantage of them, and how helpful people are in general.

- ICCS asks several subjective questions about **political opinion**. In 2009 children aged 13 and 14 were asked about their agreement with statements about gender equality in terms of rights, women in politics and men as political leaders, the labour market, and equal pay.

- Information on the **personal perception on the locality** in which respondents live is collected in the HBSC survey. Two questions are asked, the first relates to a general perception of the area and if it is a good place to live. The second is more detailed and asks for agreement about whether neighbours are friendly and trustworthy; the area is a safe to play in; and, the area is a

good place for leisure activities (different from safety experiences introduced above).These data are available in 2001/02 and 2005/06 waves, but were no longer mandatory in 2009/10.

- **Subjective material situation** information is also collected in the surveys. ESPAD and HBSC ask questions to children about how well-off they think their family is (the ESPAD item specifically anchors the question other families in the child's neighbourhood). The EQLS asks families how easy is it for their household to make ends meet.

4.6.1 Summary of subjective well-being section

313. The most commonly asked questions on subjective well-being are school life and education. Health, body image and satisfaction with the locality in which children live are the least common. Personal relationships data is most commonly collected from adults, as are health subjective life satisfaction and material well-being items. The range of ages covered in this domain is in the majority from 11 to 16 years. The only data collected from the youngest children (aged 9) reflects their perceptions about education. Information on body image, political opinion and satisfaction with locality cannot be disaggregated by migrant status.

314. It might be expected that subjective measures of well-being would be made available to complement or reflect all objective measures; for instance satisfaction with household income (or perceptions of sufficiency) might be asked alongside items requesting actual income sources and amounts. To this end, missing concepts in subjective well-being can arguably be identified by objective measures without subjective complements (e.g. "Do you have enough school items to contribute fully to all classes at school?" and not simply a count of available items). Furthermore missing concepts in the policy dimensions would require similar subjective responses: for instance, children's perceptions of safety in the home or satisfaction with access to health professionals in times of perceived need.

Table 23: A summary of subjective well-being indicators

	Aspirations	Body image	Education	School life	Health	Life satisfaction	Personal relationships	Political opinion	Satisfaction with locality	Subjective material situation
Number of indicators	4	2	9	9	2	3	6	6	2	3
Number of sources	2	1	3	4	2	3	4	1	1	3
Adult data (parents)	0	0	0	0	1	1	4	0	0	1
Child data	4	2	9	9	1	2	2	6	2	2
Child ages	13; 14; 15	11,13 and 15	9/10; 9-13; 15	10; 11,13, and 15; 15	11,13 and 15	11,13 and 15; 15/16	15; 15/16	13/14	11,13 and 15	11,13 and 15l 15/16
Migrant	3/4	0/2	9/9	5/9	1/2	2/3	6/6	0/6	0/2	2/3
Scales	0	2	8	8	1	2	5	6	2	1
Ordinal	3	0	0	0	1	1	1	0	0	2
Nominal	0	0	0	1	0	0	0	0	0	0
Binary	0	0	1	0	0	0	0	0	0	0

4.7 An over view of data and gaps in child well-being indicators

315. Overall, there are large amounts of data available on children in European and OECD countries. The European focus in many of the surveys means that European children are better covered in the data, what is yet to be addressed fully is the quality of the data that is already collected.

316. Of around 350 survey items identified covering the different dimension of child well-being around one quarter to one fifth cover socio-demographic factors (Table 24). This is due to the necessity of these types of data for interpreting the main results in the different surveys, and can offer opportunities for data harmonisation.

317. The next most populated dimension is income and deprivation, then health and risk and civic participation and time use. Income and deprivation are important indicators in the household surveys, indeed over two thirds of the questions in this domain are asked of household heads. Less is known, in general, about housing and environment and children's relationships: in both cases these provide broader contextual information necessary to understanding the environments (emotional and physical) in which child grow. In the former, adults' data is most often available, in the latter child data is the main source.

Table 24: Summary table of cross-national child well-being indicators by dimension in the EU / OECD regions

	Socio-demographics	Education and schooling	Health and risks	Income and deprivation	Housing and environment	Civic participation and time use	Relationships	Subjective perceptions and opinions
Number of indicators	81	45	56	79	19	56	17	46
N. of sources (ave.)	5	3	2	3	2	3	3	2
Adult data (parents)	31	6	6	51	17	12	4	7
Child data	50	38	50	28	2	44	13	39
Child ages	9-15	9-16	11-16	9-15	11-15	9-16	9-16	9-16
Migrant	98%	87%	59%	89%	89%	68%	71%	61%
Scales	22	24	21	21	5	36	6	35
Ordinal	14	10	24	10	2	7	7	8
Nominal	15	1	0	11	3	5	0	1
Binary	30	10	11	27	9	8	4	1

318. In terms of age, children's own responses in each dimensions data are completely missing before the age of 9. By dimensions, health and risks and housing and environment do not have age specific responses for children under the age of 11.

319. Overall, there remain important gaps in the knowledge of children's well-being, by dimension, in the area of child protection and neglect. More broadly, the cross-sectional nature of the data means little is known about the persistence of experiences (or development) in each of the dimensions.

CHAPTER 5. RECOMMENDATIONS FOR THE USE AND IMPROVEMENT OF INTERNATIONAL DATASETS ON CHILDREN

The purpose of this final chapter is to provide recommendations for the use and improvement of child well-being data. Because the business of developing and delivering international surveys of children is complex and costly, the most effective way of meeting ambitions for monitoring child well-being is to highlight what can be achieved with the available data, and recommend amendments at the margins of present surveys. Recommendations are divided into those for survey coordinators and those for researchers and policy makers interested in developing measures for monitoring child well-being.

It is critical for survey coordinators to recognise and react to the demand for data in policy and academic circles, and to encourage its use through open access, technical supports, and questionnaire development. The broader value of the surveys is, in part, through contributions to evidence-based policy. Notwithstanding the competing interests between survey development and consistency for time trends, a number of steps can be undertaken, including: consulting the research community more broadly (via internet forums for instance) in regards to questionnaire changes and developments, harmonising surveys or parts of surveys with other international or national studies, and the development of rotating modules, supplemental out-of school surveys, or longitudinal components to the studies.

For researchers and policy makers, an important challenge is to inform policies for children today, using what is essentially incomplete data on child well-being. Bad evidence is worse than no evidence, yet producing no evidence is undesirable when good evidence is possible. Indicator development requires case-by-case validity and robustness testing, and efforts should be made to appropriately 're-weight' the sample data to account for non-response in non-core items (for example, achievement scores in educational surveys). Efforts to carefully highlight which child populations are missing from the estimates are also necessary; survey coordinators should make providing detailed and 'easy-access' evidence on missing populations a priority.

5.1 Introduction

320. Undertaking and publishing international surveys of children, or households with children, is a costly, complex and time-consuming task. For these reasons, it is unlikely that many new surveys of appropriate coverage and quality will become available in the near future. The best opportunity for improving the quality and availability of data for informing policy decisions on child well-being will come from amendments to existing surveys, and indicators derived from these surveys. Moreover, in a context of limited public resources, and the fact that school can be over burdened with surveys, it is important to make the best use of available international survey data on children.

321. This chapter is organised in three substantive parts. The first looks at the available data for assessing child well-being in European and OECD countries. The second provides recommendation for improving the available data on children for cross-national monitoring, splitting recommendations most

relevant for the survey coordinators from recommendations for the calculation and use of child well-being for monitoring and policy informing. The final section of the chapter summarises the key messages for the use and improvement of international datasets on children.

5.2 Key recommendations

322. For effective cross-national comparisons, it is not only necessary to develop reliable and varied indicators, but the development of a monitoring tool for child well-being across countries is also a priority (as highlighted in a number of key meetings and publications including the Child well-being consultations, UNICEF, 2007, EC, 2008, and OECD, 2009). The 'gold standard' for such a tool would be: a valid, robust and repeatable tool that captures information across a range of well-being dimensions, the different experiences by age and for children at particular risk (e.g. migrant children, institutionalised children). Such a monitoring tool would quickly capture the data, and allowing for timely comparisons, as well as comparisons of trends over time and between dimensions; it would be represent the overarching latent concept of child well-being, but also easily breakdown to dimension and indicators-level in order to encourage considered policy / civil society responses. All of which begs the question, what is needed in order to achieve this?

323. Before discussing detailed barriers to well-being measurement and recommendations for use and improvement (for researchers and survey coordinators), a short list of priority recommendations should be introduced for the purposes of clarity. These recommendations are selected on the basis of more detailed discussion in the various sections below:

- **Concerted efforts should be made in policy and research circles to fill gaps in child well-being comparisons** in terms of both age-related indicators (children under nine are missing from survey work) and in terms of new dimensions and indicators of child well-being, not presently covered in the studies (child protection, mental health measures, and more recently civic participation). There is a role for all parties in this process; from the demand-side in terms of policymakers and researchers, and the supply-side from funders and survey coordinators. Although some series data is available, survey data is needed for detailed analysis and recommendations. Some survey data is already available to do this in Europe (ESS and ESPAD), for non-European countries new data (possibly via amendments to present collections) are needed.

- **Communication between survey coordinators and survey users can be improved.** For instance, surveys should be encouraged to communicate their proposed changes wave-on-wave, and consult users (through open forums or bilateral discussions). Policymakers and researchers should be encouraged to provide evidence in support of claims, with respect to the stated purpose and goal of the survey in question. This will help with planning for monitoring and research purpose, and enable survey coordinators to reflect on priorities.

- **The use of equality indicators and social gradient indicators in monitoring child well-being are necessary.** Policymakers and researchers need to use equity indicators when assessing outcomes from public services. Researchers should use available data to explore inequality across more indicators, and surveys should be encouraged to produce scalable items in new questionnaires and / or amendments to present items. Although not yet available, currently the HBSC Family Affluence Scale is being developed and updated with new items for use in 2013/2014 survey – the adapted FAS should strive to capture a greater degree of variation in social gradient than the original measure.

- **All new child well-being indicators derived from secondary analysis of surveys should be validated, and if necessary, treated for non-response bias.** Researchers should undertake post hoc weighting adjustments to indictors with bias in non-response (see 5.5.2); when more complex multivariate analysis is undertaken, to ensure representative estimates to fill gaps are needed. Policymakers should ensure that validation tests and appropriate adjustments have been undertaken before allowing the findings of secondary analysis to directly inform policy decisions. Survey coordinators should provide guidance, when relevant to their specific survey data, to cover methods for validation or treatment of non-response bias in their surveys.

- **All indicators and analysis derived from present school/household surveys should clearly show which children are missing from the analysis, and effort should be made to fill those gaps** (see 5.5.1). To facilitate this, in the first instance, survey coordinators should report on missing populations (with the help national survey managers), or consider supplemental out-of school surveys.

- **Efforts should be made to harmonise data collections between the major international surveys.** This will facilitate analysis on interactions between various well-being experiences and provide new evidence for policy presently not available. Where national sets have been quality checked, national efforts to harmonise international evidence should also be undertaken. Funding bodies or governmental bodies, that work with more than one survey, can be involved in these efforts.

- **Notwithstanding financial constraints, funders and governments should be prepared to support a more comprehensive international survey of children** (by child well-being content and age) if present surveys cannot adapt to meet the necessary requirements for properly informing policies that aspire to cross-sectoral integrated interventions for *all children*.

- **If efforts to improve the knowledge base for comparing child well-being are to be meaningful, it is essential that policymakers respond to the availability of new data and evidence with advances in monitoring and policy development.** In doing this, it is necessary to consider which indicators are most appropriate for informing policies for childhood in different age-groups, and in different at-risk categories, and use them in target-setting and the retrospective evaluation of both new *and* on-going policies for children.

5.3 What available data can be used to assess the well-being of children across countries?

324. Chapter 4 has clearly shown the wealth of available data, not only for assessing child well-being across EU/OECD countries, but also for detailed assessment of education, health, and material well-being considerations. As a group, the European nations have more survey data available for the purposes of monitoring than the OECD set (which, apart from PISA, is not presently undertaking membership-wide micro-level surveys in this area).

325. Many indicators have been developed to represent aspects of child well-being in recent years (see Table 7). In most cases these have then been aggregated into dimensions or overall measures to represent the parts, or the whole, of child well-being. What is clear, is that much more data is available in cross-national surveys to contribute to these works if considerations such as country coverage, age coverage, and timeliness of data were to be less of an issue.

326. Looking only at the most recent data (past 5 years), over 380 separate child well-being items have been identified in eight common dimensions, over 42% of which are collected in scale form (both discrete and continuous) – allowing for complex forms of analysis and inequality indicators to be

developed. In each dimension, at least one in ten of the indicators are available for a complete breakdown by age (the lowest proportion being in health and risk at 11% - although the children themselves are not the respondents). In around 70% of cases overall, information can also be reported for migrant children separately.

5.3.1 So what should policy makers demand?

327. At the moment, in the comparative field, developing a monitoring tool of 'gold standard' is not possible. This is due to several major challenges for the development of indicators of child well-being – and so policies for child well-being – which exist because of limitation in the international surveys and research undertaken with those surveys. Policy makers should be particularly aware of these limitations in present comparative research on child well-being, and demand improvement in the available data and research in order to fill these gaps.

328. The following recommendations reflect upon major 'structural' challenges, and should be a starting point for policymakers looking to achieve a 'gold standard' in evidence based policy. More detailed recommendations on how these changes might be achieved via the work of survey coordinators, or researchers, is presented in section 6.3.

- *Without present surveys being expanded, new surveys will be needed to cover unmeasured aspects of child well-being and hidden populations.* Cross-national comparisons of child well-being have been largely data driven (both inputs and outcomes), and in the case of comparison of child well-being outcomes data is mostly taken from surveys of adolescent children. Using a limited set of topic-specific school surveys means both content and coverage are below optimal levels from a child well-being perspective. For instance there is no detailed information on children's experiences of neglect, interaction with services, professionals or other adults outside of school, relationships with extended family members, detailed time-use, or various life preferences (including consumption, leisure, and learning). Moreover, children who are out of school (many Roma children in Europe for instance), not in mainstream school settings (see annex 8), are absent or excluded are not captured by school surveys – despite being the children with high risks of low well-being.

- *Age specific indicators and monitoring tools, and in turn, policy evaluations of age specific interventions, are needed for a closer look at experiences in the different stages of childhood.* Data for the youngest children (under 6) are collected only in service settings (e.g. hospitals) via series data and as such do not fully represent the various domains of their well-being, and cannot be effectively disaggregated for within country variation or distributional analysis. Moreover, this imbalance in measurement of well-being means age-specific monitoring of well-being is not possible, and so evaluation of age-specific policies – such as maternity polices, or policies to support successful school to work transitions – is confounded in the analysis of aggregate outcome statistics. In short, with present cross-national data, many direct links between policy effort and outcomes cannot be made.

- *Internationally comparable survey data needs to be consistently collected on early childhood.* The lack of internationally comparable survey-derived and representative data on children under the age of nine raises the possibility of a new international survey of pre-primary school children. The present lack of survey data makes it very difficult to assess inequalities and social gradients in early childhood internationally. Possibilities for filling these survey data gaps without undertaking a new survey include: harmonising available national data of key measures from national surveys, or promotion/application of harmonised birth cohort studies in more countries. If new surveys were to be undertaken, an Early Development Instrument (EDI) approach would

be preferable (EDI looks at developmental outcomes, has a proven method of agency in the collection of child data, and has a strong applicability at the local level where many family services are delivered – see for instance, examples in Canada or Australia).

- *Monitoring tools and indicators, so far, do not capture the experiences of different children where possible, including disaggregation by sex, migrant status and at-risk groups (poor children for example).* While a number of national studies, and smaller country comparisons, have been able to focus their analysis on children by sex, migrant status or vulnerabilities, the main international comparisons in advanced economies have focussed on the 'average' child. This limits the ability of these tools to inform policy development, and interventions, for the purposes of improving child well-being as inequality in the experiences of different outcomes by these groups are not identified in standard 'average child' models.

- *The age of survey data can range from 2 to 5 years at time of publication, or public access, because of the logistical constraints of delivering international surveys,* because of embargoes, or because of the needs of multi-dimensional tools to align survey results delivered in different years. More efforts could be made to prioritise the timeliness of all or some data, based upon the selection of key indicators. Real time measurement of child well-being outcomes are essential for effective policy response, particularly in view of experiences such as the Great Recession, and so steps should be made to begin regular collection of data required to monitor child well-being in advanced economies in a timely and systematic manner. Agreements between survey coordinators, funding organisations, and organisations with international monitoring plans could be undertaken.

- *There is little international evidence on how well-being measures interact to allow for a prioritisation of well-being indicators to be undertaken that is sensitive to ages and national setting.* To do this, more comprehensive datasets of well-being would need to be made available through new collections or harmonisation. Failing that, where national surveys are available, these types of interactions are calculable country-by-country. Evidence and recommendations on how to appropriately set selection criteria for indicators development and use, would also provide much needed guidance in regards to when indicators may have outlived their usefulness in comparisons, as technology, society and childhood changes.

- *The methods of representative child participation in the development cross-national research on child well-being need to be addressed.* Child participation is a central tenet of the United Nations Convention on the Rights of the Child, and yet – in the most active sense – it remains underutilised in the comparative research field (with the exception of child surveys). However, cross-national representation of child participants in research is a huge task, equivalent to another international survey of children of all ages. Moreover older, more able and less vulnerable children are more likely to be active participants when appropriate national coverage is achieved. To overcome these constraints on effective child participation, research should be commissioned on the methods and application of child participation in international research.

- *Evidence-based recommendations on the aspects, and extent, of child-centred indicator development and definitions are limited*. Any monitoring tool should critically reflect on how 'child-centred' the indicators used for comparisons are and whether child participation is required and how such participation should be undertaken. For instance, is it sufficient to use child income poverty measures that account for child populations instead of numbers of households with children? Or should additional steps, such as identifying how much money is actually spent on the children (their necessities / luxuries) be undertaken.

- *Attempts at new surveys in the field of child well-being are stifled by underfunding, and by the already high demand on schools to undertake surveys.* To facilitate efforts to fill gaps in the knowledge, the international policy and research community should support new survey efforts. To do this at lower costs, overall and on responding schools, the funding bodies involved should encourage cooperation between existing surveys and new ones.

329. To overcome many of these major challenges, policy makers, and particularly governments that fund surveys and research, should consider facilitating change through targeted investment in present surveys and further exploratory analysis.

5.3.2 What data is available and how should it be used?

330. Beyond the methodological advances required to fully inform policies for children, there is a need to better utilise the available data on childhood to inform policies for children. This, in particular, is a role for researchers. The main points on what data is available, and how it should be used, are as follows:

- *Of all available survey data, more information is collected on socio-demographic factors than any other separate 'dimension' of children's lives*. The wide coverage of socio-demographic factors is due to their role as explanatory factors (or controls) in all topics for analysis. More could be made of this data, first as a point for the harmonisation of survey results, and second for complementing wealth or income indicators for analysis of social gradients and inequality. So far, the major cross national comparisons of child well-being have not presented variations by family type (although efforts to understand the role of family form on child well-being have been undertaken – see OECD 2009 and Bradshaw *et al*, 2009).

- *Following socio-demographics, income and deprivation are the most popularly informed aspects of well-being.* Income and deprivation indicators, above all others, have been integral to policy goals for child and family living standards in the EU and beyond, for a number of years now, and their popularity is therefore not a surprise. That said, the use of income indicators in the analysis of social gradients in outcomes of other forms of well-being is underutilised.

- *Health and risks, and civic participation and time use are also well populated***.** HBSC and ESPAD are responsible for the former, and ICCS contributes to civic participation data as well as CIVED (now non-operational), whereas time-use data is spread across a range of surveys. HBSC data on health and risks has been well used in efficiency (averages) and equity (distributional) measures of well-being in cross-national studies. Civic participation and time use data is less well-used and should be incorporated into studies (as surveys such as ICCS become available) to represent children's autonomy (locus of control), which has shown to be important factor for driving child well-being in national studies (see The Children's Society, 2012).

- *The education dimension is one of the least populated ones, with 45 available survey items*, one less than the subjective dimension. This may in part be due similarities in how this dimension is conceptualised and measured as opposed to reflecting its importance (via achievement or performance scores, or attainment levels). Despite the limited availability of items, education indicators, and particularly achievement scores, can lend themselves to complex analysis of distributions, and multilevel models. Because of their broad background questionnaires – including parental questionnaires - PISA, PIRLS and TIMSS provide a unique opportunity to analyse how well-being in childhood contributes to educational achievement.

- *The least is known, in general, about housing and environment and children's relationships*, despite the fact that these types of measures will affect the ability of policy to influence well-

being outcomes. In both cases these provide broader contextual information necessary to understanding the environments (emotional and physical) in which children grow-up. In order to encourage a greater collection of these types of indicators, and to understand how these contexts moderate or mediate variations in well-being across countries, more analysis of those that are available should be undertaken by researchers.

- *A review of available data also uncovered many more available subjective well-being indicators than previously used.* Subjective indicators are available for aspirations, body image, education, school life, health, life satisfaction, personal relationships, political opinion, satisfaction with locality , and subjective material situation. In the vast majority of cases these are available in scale form, meaning not only can most dimension of well-being be recreated from the subjective perspective, but inequality measures (using short discrete scales) can be produced (see UNICEF, 2010 for examples).

- *In terms of age, children's own responses in each dimension are completely missing before the age of nine.* Household surveys can be used to breakdown age-related experiences by dimension, although this information is provided by parents or guardians. By dimension, health and risks and housing and environment do not have age specific responses for children under the age of 11 because PIRLS and TIMSS do not ask questions relevant to these dimensions of child well-being.

- *There are few indicators available on parenting practises.* For many children, parents are important agents in achieving well-being including, social, emotional and material support. More research – taking advantage of data from parental questionnaires and child responses to questions on family – could cover parenting practises, and policies that promote – or are conducive to – good parenting practises. HBSC cover parental practises in optional packages, however less than one-third of countries are taking up the opportunity to ask these questions (14 in total in 2009/10).

- *All analysis should provide clear recommendations for policy development, and if this is not possible, explain what would be needed in order to do this effectively.* Present attempts to compare child well-being internationally often stopped short of answering questions such as: why do well-being outcomes vary as they do between countries?; and how should policy react in order to improve child well-being holistically? An important development for the researchers in the field is to provide policy recommendations on the basis of the available evidence, whilst appreciating the competing and complementary nature of both social policies and CWB outcomes as measured in the frameworks. More could be made of available micro-simulation tools, and upcoming research could explore relative benefits of service versus cash interventions for families with children (OECD will undertake work in this area in 2012).

- *In both cost benefit analysis of interventions, and suggesting indicators for children, an appreciation of the short- and long-term goals is needed.* Analysis of the relative efficiencies of various policy interventions for children at the international level can do more to recommend 'what works' for child well-being in terms of both short and long run returns on interventions in order to better inform policy priorities and public expectations. Associated, is *the need to recognise the elasticity of child well-being indicators in the short- and long-term in relation to social or economic shocks* (depreciation), for the purposes of prevention and reflection on policy choices in the short and medium term.

331. There are a number of avenues for making better use of available data for expanding indicator coverage and dimensions of well-being operationalised in cross-national comparisons. For instance, a broader set of dimensions of well-being could include indicators that represented: social and emotional

development of children (social skills such as relationship skills); civic and social participation of children; and child protection, are necessary to capture better child well-being holistically. The lack of these dimensions cannot be totally blamed on data availability, but rather on a desire to maximise country coverage in these studies. For instance measures of mental health are available in ESPAD, and measures of civic participation are included in ICCS.

5.4 How can the existing data for cross-national monitoring of child well-being be improved?

332. Using the examples of best practise discussed as part of Chapters 2 and 3, and available data in Chapter 4, this section recommends ways to improve the use of available data collected in each of the international surveys for children for informing policies, and monitoring the well-being of children. A number of developments specific to indicator development are also needed; these include indicators of distributions and inequalities, and indicators for preventative interventions. The main points are as follows:

- As a priority, include measures of inequality indicators in attempts to monitor child well-being cross-nationally. Equality in childhood is important for all children's present and future well-being. The equality of outcomes in a child population is a measure of equity, *as well as* efficiency of public services. Child well-being inequality measures can also be used to inform and monitor policies that promote equal opportunities (equity) and outcomes (equality). Both equity and equality for children across developed countries is hindered, to varied extents, by the transmission of inequalities across generations (OECD, 2009). Evidence-based knowledge of the way inequalities transmit and impact the child's development is also needed.

- Comparative indicators of child well-being are yet to be produced that account for differences in the social gradient. This is mainly due to the fact that most child surveys do not include data on household incomes. To achieve this goal, surveys should be encouraged to collect data on household's income or its proxies in present international child surveys, as well as data on the family's material and non-material assets (educational background, time availability) or other data used to identify the social gradient, such as employment situation of parents.

- To complement child well-being outcomes in the datasets, contextual data on families and the communities in which children live are necessary. To make sense of how interventions for CWB result in better outcomes for children, it is important to acknowledge the settings in which these interventions occur. For instance, family settings, school settings, or broader community settings can facilitate or inhibit a child's access to essential services (for example via parent's free-time, teacher pupil ratios, or access to public transport).

- So far, there are few child well-being items in surveys that factor in time, or persistence, of childhood experiences. The cross-sectional nature of most data collections means almost nothing is known about the persistence of, or the longer term outcomes from, experiences in childhood of each of these dimensions. Few cross-national indicators can factor in time effects. Stability in experiences, time and timing of childhood experiences, will influence the child well-being via entrenched experiences, or habituation. For policy at national level, two countries with similar poverty rates may have very different experiences of persistent poverty, and different populations who live in poverty at different times. Together these differences will have a profound effect on the true impact of child poverty, as well as change the way in which policymakers should respond to the challenge. Where possible, research should attempt to capture experiences of time and timing in outcomes measures.

- Disaggregated information by socio-demography and geography are needed. Many child well-being indicators reported at the international level are not systematically disaggregated by socio-

demography and geography. This is possible in many cases as data is available, and would provide basic information about the disparities in well-being, and in turn enrich policy recommendations.

- For effective evidence-based policy, both outcome and preventative indicators are needed. Identifying which indicators are good preventative indicators (perhaps for example, truancy at age 12 for educational achievement at 15) are necessary to design policies of prevention, and as such are an important step in improving the knowledge base.

5.4.1 Recommendations for survey coordinators

333. It is critical for survey coordinators to recognise and react to the demand for data in policy and academic circles, and to encourage its use through open access and supports. To this end, the broader value of the surveys can be recognised through contributions to evidence-based policy, and in turn recognition of, and financial support for, the surveys can be improved.

334. A number of steps can be undertaken, including:

- Efforts should be made to improve communication between survey management groups, and with cross-national organisations and research groups working in the policy field (OECD, EC, etc.). Consulting the research community more broadly (via internet forums for instance) in regards to questionnaire changes and developments, would be useful for surveys coordinators as they would learn how their data is being exploited in its various ways, and the needs and interests of a broader range of users. Better interaction might also include allowing some groups with earlier/quicker access to data in the cases where reporting is time sensitive (but indicators are not headline ones for the surveys – this may the case for child well-being monitoring purposes in some cases). Moreover, good communication will enable researchers / monitors to plan future evaluations.

- Efforts could also be made to improve communication between survey coordinators themselves, for reason of harmonisation, or to develop complementarities in conceptual coverage. Examples exist, for instance, there is ongoing dialogue between HBSC and ESPAD (regular conference calls between dedicated members from each study) to discuss survey items in common, survey timing, timing of reporting, sharing ideas.

- Surveys should strive to achieve basic requirements for time trends analysis, and consider ways to capture longitudinal experiences of schools, if not households or children themselves. With the exception of PISA longitudinal (undertaken in 4 OECD countries – Australia, Canada, the Czech Republic, and Switzerland), longitudinal analysis is not possible. And despite much good work on cross-sectional comparisons of child well-being, only recently has the availability of new data – particularly from ongoing international child surveys – has recently made time trends analysis possible. How changes in child well-being measures had trended in recent years will not only prove useful for identifying priorities in child well-being measurement, but would also prove useful for the validation of child well-being statistics. In order to do this consistency across key indicators and contextual factors is needed across waves. HBSC is a good recent example of an international survey that has made efforts to develop trends over time (their questionnaire has been unchanged over three waves – 2001/02, 2005/06 and 2009/10). Five surveys in total explicitly mention developing time trends as a goal of the survey; consistency in questionnaire design (whole or in part) is necessary to do this.

- Much more could be done in terms of data linkage and harmonisation of national and international datasets. Data linkage, and opportunities to harmonise data or parts of collections, in new and existing international surveys is key to making the most of efforts to inform good policies for children. One way harmonisation could be achieved, is through linking certain items in international child surveys to allow estimations to be derived for unobserved variables. This would make possible more detailed analysis of the interaction between observed and estimated child well-being outcomes (see Box 2).

- When data is released for public use, countries with significantly lower than average enrolments should be flagged for cautious interpretation. Chapter 3 has presented concerns about bias in school surveys driven by enrolment rates. To aid researchers, survey coordinators should make providing detailed and 'easy-access' evidence on all key hidden populations a priority (see 5.5.1 for suggestions on how this might be done). Advice, or guidelines, on how to develop weighting systems to account for non-response on key indicators would also be welcome (see 5.5.2 for an example).

Box 2. Harmonisation: Methods for the statistical matching of survey data

Given the uni-dimensional nature of the main child surveys, and the widespread acknowledgement that child well-being should be covered by multiple measureable life dimensions (education, health and income for instance) it is necessary for researchers to be able to match results and indicators of child well-being at the micro-level to understand how these variables interact. Beyond the analytical goals, data-matching would also provide new opportunities to test the robustness of results between similar surveys (risk behaviours, or education results), or produce better estimates to counter non-response due to survey fatigue or language barriers.

The manner in which datasets can be matched is subject to much complex statistical discussion, beyond the realms of the work undertaken here, but it is possible to discuss the very practical requirements of harmonization that would allow the more complex steps of statistical matching (such as Propensity Score Matching or the imputation of data across surveys) to be undertaken.

Matching child well-being datasets means harmonizing items across different surveys. Harmonized items are then used to identify comparable cases or subsets in each survey, which in turn can be used to match results for different subsets across surveys, or to produce estimates for imputation across surveys by case (for instance to impute literacy measures from PISA into the same dataset alongside estimates of health outcomes available in HBSC, or civic participation from ICCS, or literacy in TIMSS or PIRLS and so on).

In order to match data at the micro-level, for the purpose of comparing separate child well-being survey results here, some considerations are necessary (see Steiner and Cook [undated] and D'Orazio [2011] for more details on practical steps by matching technique):

- Two separate surveys with the same target population (including target populations in surveys designed to produce reliable estimates at sub-populations of children by age or gender) are needed.

- The two surveys should have items that collect the same information, to provide a comparable set of variables from which to match the results for respondents to allow for unbiased comparisons of results across surveys, or for the identification of 'donor' cases from one survey sample to provide estimates for unobserved child well-being outcomes for 'recipient' cases in another.

- For the imputation of survey data between surveys, matching variables should be good predictors of the two or more independent variables that will be matched from the different sources; and the final matching variables should selected from the subsets of matching variables that are predictors of the independent variables in regression tests undertaken in the two separate surveys.

- For imputation, the two separate surveys should have two variables that are assumed to be 'conditionally independent'; meaning the knowledge of the level of one, and it's association to the independent matching variables, should not help you predict the level of the other.

- Results of any matching should be properly evaluated for robustness where possible (this can be done using alternative approaches to assessing the association between variables derived from different surveys, such as the evaluation of uncertainty – see D'Orazio, 2011).

Based on these steps for statistical matching, harmonization of child surveys requires a sufficient number of key variables relevant to the work of all surveys to be matched. Detailed examples of which specific indicators might be harmonized, and necessary adaptations to questionnaires, are discussed in section 5.5.3.

- All of the surveys are good at exploring their subject matter (education, health behaviours, and so on), but despite best efforts, are limited by space and time when it comes to covering a full range of contextual factors that might to help explain their findings. This is likely to become an increasingly important issue over time as the role of 'softer' measures in explaining what works in these area (such as parenting practises, family relationships and subjective well-being) gathers

momentum. Almost all of the surveys mention policy informing as a key purpose of the work – efforts to capture new relevant information and contextual information, to stay at the forefront of research developments, are needed.

- Survey coordinators should react to demand for inequality measures in child well-being. Although the majority of data collected by surveys are collected as scale variables, these are often discrete scales, and are underutilised for analysis of distributions and inequality. Surveys should be encouraged to explore opportunities for developing items that contribute to both average and distributional measures, particularly with new items, and particularly with items where underlying inequality (and so the relative nature of the of the outcome) are most relevant. Both ESS and TIMSS explicitly state that equality measurement is a purpose of the data collection, and for this continuous scale variables are needed.

- Efforts could be made to capture indicators for populating dimensions presently not available, particularly in surveys covering non-EU countries. In part this will be keeping up with recent research interests. Mental health measures, for instance, are available in ESPAD which asks children about mental health symptoms, such as losing their appetite, feeling depressed, pressured or sad, and not be able to work (school work) or having difficulty concentrating. A second ESPAD question directly asks children how often they have self-harmed or attempted suicide. Yet despite recent OECD research covering various issues such as mental health, asthma, and disability, no survey covers these types of data OECD wide. If existing surveys wish to cover these gaps, the use of obligatory rotating modules is one way to achieve this.

- Advice from household survey coordinators is required for appropriate weighting techniques for child populations in these surveys, particularly breakdowns by age groups. Household surveys are important for covering the age spectrum, because of the limitation of child surveys. Good practice guidelines – beyond household weights by number of children – are needed.

- Population weights which do not inflate sample numbers (N's) (i.e. weights represented as ratios) should be applied to facilitate more accurate confidence margins, and to avoid inadvertent Type 1 errors in results. Considering that many of the child well-being indicators are developed from non-core items, with non-response issues, non-N inflating weights use sample N's for calculating standard errors, and so err on the side of caution in regards to confidence interval and significances.

- More time sensitive measures, such as length of time living in a sole parent household, are needed to inform the true influence of these factors on child well-being. Good examples of this type of measure are found in TIMSS and PIRLS where there are not only questions about speaking the test language, but also questions about the time spent speaking the language at home, and more specifically the time spent speaking the language with adults in the home (PIRLS). In TIMSS, PIRLS and PISA migrant children are asked their age when they entered the country, which provides additional detail needed to analyse the time the child has had to integrate into the country of test. In EQLS, questions about household debt, rent or mortgages and utility bills, a time limit is also given (experiences in the past 12 months).

- Efforts to contextualise data collected in other ways can be used to explore key indicators. For instance in PISA 2000, children were asked how homework was contributing to class work (counted as part of marks) their interest in homework, where they did it, how much they did, and with what distractions (for instance, if the TV was on). The questions about homework in TIMSS simply ask how often children get mathematics and science homework, and the length of time it takes to complete it (no reference to the amount of homework). This type of focus can be rotated

into background questions, like the literacy focus in PISA is at present, if restrictions based on other issues (such as time trends analysis) allows.

- More experimental methods could be used to get at validity issues in international surveys, particularly around sensitive items (such as risk behaviours, or parental and peer relationships). A good example from ESPAD is the inclusion of 'relevin' in a list of drugs used by the respondent (sedatives, amphetamines, LSD, crack, cocaine, heroin, mushrooms, GHB and steroids are also included). Relevin is a dummy drug used to test children's honesty in answering these questions. Analysis of relevin responses can be used to adjust the findings of real, but sensitive, items included in the survey.

Box 3. The value-added (or not) of interviewing children directly

This report has looked at international surveys in Europe and the OECD that can be used for informing child well-being. In the context of child participation discussions, and with an eye on the potential for future surveys to be developed, a brief discussion about the added-value of interviewing children directly is relevant.

Despite discussion on 'whether to interview children', or use child-reported data for child well-being comparisons, being broadly based on: issues of accuracy of responses (the assumption that children may not know the important 'facts' for objective responses, or are less reliable than adults when providing 'subjective' responses); and, the importance of child participation in data collection process, there are sound pragmatic reasons why the surveys that are in the field interview who they do when assessing life experiences or well-being.

In the case of child surveys, clearly PISA would not be able to assess the child's true education ability unless the child themselves responded to the tests. Moreover, when assessing health behaviors or risks, it is unlikely that parents of adolescents would be able to give a full and frank account of their children's experiences. In contrast, household surveys that concentrate on living standards and household incomes tend not to interview children as an adult head of household is already providing information on their behalf, and the children are not considered well-placed to provide full information on household living conditions and financial circumstances.

In the field of child well-being, interviewing children directly is important to achieving the child's rights, and ensuring that their voices are heard (UNCRC, 1990). To this end, it is also important that children participate, where feasibly possible, in the development of the topics and items used to operationalise and evaluate their experiences of their well-being. The critical issue for child participation is to involve *all* children in the design and collection of surveys, and to consider methods in identification, sampling and survey delivery that will facilitate this (out-of-school interviews, online questionnaires, or though observation).

Having highlighted the need to interview and involve children, it is also important to acknowledge the fact that not all children can participate in surveys, and to appreciate the need to go to other sources to help represent children's well-being (for instance when they are too young, or do not have the ability or knowledge - for various reasons - to provide answers to questions deemed critical to informing and ensuring their well-being). For example children would normally not be able to report details of family incomes and outgoings (as a driver of various well-being outcomes, and an important policy instrument). Moreover, in order to capture the well-being experiences of infants longitudinal surveys of children's well-being, the child's parents or guardians will need to represent the children in that survey.

In order to collect reliable and accurate child well-being data, that helps children achieve their rights, interviewing children is absolutely necessary. In particular in circumstances where information collected attempts to reflect personal experiences, needs or aspirations of children. To ensure that all children benefit from these surveys, new methods to identify and sample hard-to-reach / at-risk child populations are needed. However, there remains an important role for parents and guardians (including service providers, such as school principals) in representing some children, and in providing the contextual information of the 'adult-world' that might help explain differences in children's well-being, as well as develop efficient and effective policies for improving child well-being.

5.4.2 Recommendations for the use of child well-being indicators

335. The key message for researchers in the field of child well-being measurement is that indicator development requires case-by-case validity and robustness testing, and that all efforts should be made to appropriately reweight the sample data to account for non-response in non-core items (for instance items not directly linked to the core measures in the surveys, such as achievement scores in educational surveys). The main messages include:

- Efforts to carefully highlight which child populations are missing from the estimates should be undertaken for all indicators derived from child and household surveys.

- For sensitive measures, efforts to test for validity – including via the use of available 'dummy' questions – should be undertaken as a matter of priority.

- All indicators should be tested for non-response bias using weighted and unweighted data (see examples in Chapter 3).

- In the case of non-response, evidence of bias should be adjusted for using post-hoc weighting techniques (see 5.5.2).

- For complex analysis, tests for bias should be made on populations excluded from the models. Results of these tests should be reported alongside model findings, or where possible techniques to estimate likely responses (Maximum Likelihood Estimation) should be undertaken to maximise inclusion.

- Cautions should be provided in terms of non-response as associated to type. Including the differences between refused data, and hard to collect data. This may inform improvements to items in future survey waves.

- Researchers should explore new analysis options outlined in section 5.3. In particular efforts to develop inequality measures, expand dimensions coverage (from available data – subjective well-being, time use and civic participation), and incorporate time and timing dimensions to indicator development.

- Researchers should analyse child well-being indicators in the context of broader economic contexts and policy efforts. Understanding the economic and political context provides insight into the complementary and competing goals that will define the boundaries of future policy interventions. Family policies are not the only route by which to achieve child well-being, and similar family policies might function different in different national economic settings. An appreciation of the economic and policy context is likely to lead to more suitable, and effective, policy recommendations from the field.

5.5 The way forward: three key issues

5.5.1 The Why and How of surveying hidden populations

336. An important part of fully understanding levels of child well-being in developed countries, and which indicators best represent the experiences of children in different circumstances, is representing 'hidden populations'. Many of the indicators of child well-being in developed countries are derived from survey data, and the term 'hidden populations' refers to groups in societies that are not included in official statistics on childhood. The reason for this exclusion is mainly due to the ways in which the surveys are

collected (school surveys that do not capture children not in mainstream schools, or households surveys not including families that are homeless or living in precarious conditions), or the geographical areas which are not included in the surveys (in some cases, countries are represented only in part like Germany and the United Kingdom in HBSC).

337. Using censuses, and other collections methods to study at-risk populations, can provide a picture of acute risks and different policy challenges in the Roma population in Europe, and indigenous populations in countries such as Canada and Australia. A number of factors that are associated to their exclusion from data collections drive these risks, and can include: insecure or unofficial housing, living in remote geographical locations, demographic trends over time (family structure, large families, life expectancy of non-dependants), low levels of education (or engagement with learning and training opportunities) and lack of formal work experience/engagement in the family. Because many of these hidden populations, questions are raised about the representativeness of indicators derived from international surveys, and importantly, ways in which surveys could capture these populations.

338. Although appropriate cost benefit analysis should be undertaken beforehand, several steps could be undertaken – consistently across all countries – to improve the representation of hidden populations in international child and household surveys:

- Surveys of children could stop excluding schools or pupils on the basis of educational needs, or be more selective with exclusions, by including the school or children in all or part of the survey.

- Child surveys could work with mainstream schools to identify absentees, or excluded pupils, and collect socio-demographic data on those groups, or approach those groups for responses to the questionnaires (whilst recording their status).

- Surveys could report population data on all excluded groups in order for researchers to explore the extent of bias in results by country, and make post-hoc adjustments where possible.

- The use of available census data, or unique targeted censuses / surveys to compliment the main survey results could be undertaken every few waves, and harmonized with the main data collection.

- Specialised sampling methods, such as Respondent Driven Sampling which uses social networks to identify hard to reach populations, could be applied at the margins of school or household surveys (or as a complement to unique surveys or requests to schools to find absentees). Respondents identified in this manner could then be included or excluded from analysis on a case-by-case basis; combining results with weights derived from available census data could produce useful estimates of child well-being in these hidden populations.

339. In the meantime, it is important for users of international surveys to actively acknowledge the limitations in representativeness of survey data resulting from hidden populations; and make data on this issue available where possible.

5.5.2 Using weights to adjust for non-response in indicator development

340. Building indicators of child well-being often means selecting items from the questionnaires of child surveys or household surveys, weighting the dataset using the weights provided, and running basic descriptive analysis at the national level. Given that in many cases, where the items are placed in the questionnaire (which can change wave to wave) and how they are asked will affect the levels of non-

response level county-by-country, adjustments for differences in non-response is essential for comparing indicators developed at a national level between countries and over time.

341. What is more, non-respondents by country may not always be a random sample, as certain questions in surveys may be more or less attractive to certain groups, and illicit very different response rates (Annex 4 A.2 to 5 for examples across four waves of PISA). Moreover, achievement does associate with overall responses (Annex 4 A.1, again for PISA). The result of such bias is that child well-being indicators may not represent the populations included in the surveys, and therefore weights designed by the survey coordinators to create population estimates recreate this non-response bias.

342. Adjustment for non-response in single indicator development can be achieved through the following steps:

- Using multiple selected socio-demographic indicators for bias (in binary or ordinal form – compute if necessary), such as sex, family type, wealth and education levels match proportions of the subgroups in the respondent population (RP) on the item for analysis with the overall sample population (SP).

- Calculate SP/RP to produce adjustment weights. An example in Table 25, for underrepresentation of girls with low FAS scores on a given item, the adjustment weight of 1.15 (inflating responses by 15% overall) is calculated as SP/RP or 15 / 13 = 1.15.

- To refine the selection of socio-demographic indicators, the odds of non-response by these factors can be identified using logistic regression analysis.

- If standard survey weights are to be used, multiply by the adjustment weights.

Table 25. An example of adjustment weights by sex and family affluence

Sex	FAS*	Sample pop (SP)	Respondent group (RP)	Adjusted weights (SP/RP)
Girls	Low	15	13	1.15
	Medium	16	16	1.00
	High	19	21	0.90
Boys	Low	12	14	0.86
	Medium	12	16	0.75
	High	26	20	1.30
	Total	100	100	…

Note: FAS refers to the Family Affluence Scale produced by HBSC. Figures are examples only.

343. For non-response in the overall survey sample, repeat the first step matching population figures (from a census) to sample respondents.

344. Despite this method providing a better representation of population estimate than non-adjusted indicators, there are some caveats regarding adjusting for non response bias. The first and most important is that bias can be reduced but not eliminated. Using sophisticated multivariate predictor models is likely to identify the odds of being a non-respondent would throw up more factors than those which could realistically be adjusted for using this method. IT is important to acknowledge in these cases whether more complex forms of bias have a theoretical link to the subject of the indicator, and as such represent an important exclusion.

345. A second caveat is that this method is only applicable for single indicator weights, and not applicable in multivariate models data because of the 'case-by-case' adjustments required. For treatment of non-response before multivariate analysis, methods such as maximum likelihood estimates can be applied to 'fill-in' values for non-respondents (for example, as is done in for composite calculations PISA).

5.5.3 Harmonising data collections

346. The practicalities for harmonisation between child surveys to work have been outlined in Box 2, this section now goes on to suggest specific indicators which could be harmonised across surveys, and presents a rational for their selection.

347. Based on the ways of harmonising data, and the limited spaces in compulsory section of international surveys, a number of selection criteria for identifying possible candidates for harmonisation can be identified. First, items should be common interest items, insofar as they have an independent value to each survey in the standard form they would take across surveys. Second, and related, the items should be reasonably considered standard controls in tests of, or independent predictors of, variation in the main child well-being outcome of the survey in question (whether it be education, health, living conditions, civic participation, risk behaviours, or other). Third, items should suffer as little as possible from non-response, to maximise usability, and therefore should be simple items that all children are likely to answer without difficulty (and possibly placed towards the front of each questionnaire). Fourth, the number of items should be sufficiently limited to be actionable, but also enough to ensure that a subset of significant predictors of two independent child well-being outcomes across surveys is produced in order to accurately match data.

5.5.3.1 Harmonising socio-demographic measures

348. Based on these criteria, the first suggestion is to harmonise the surveys using data that is already used to collect socio-demographic information (an option raised in Chapter 4). Table 26 provides information on specific items that could be used to harmonise child surveys, and includes: age, sex, family form, and migrant status.

349. Each item is explained in detail, and taken either directly from one of the surveys, or adapted from examples across surveys (shown in the two most right-hand column of the Table).[27] The important consideration when assessing whether an item in a survey needs adapting is whether it produces different output, on a different scale, that cannot be standardised across surveys without losing sufficient variation in responses. Where questions in surveys have the potential to produce very different output (sex and age data should be matched easily), a new item may need to be agreed across surveys.

[27] In both tables 26 and 27 some simplified items, reflecting questions asked in more than one item across surveys, has been suggested. The reason for this is to meet selection criteria outlined above, particularly in terms of applicability, simplicity and parsimony.

Table 26: Suggested socio-demographic items for harmonisation across surveys

Matching indicator	Subject or question	Item detail	Survey use similar item	Survey use exact item
Age	When were you born?	Month and year	All surveys	PIRLS, TIMSS, HBSC
Sex	Are you a boy or a girl?	Tick box: Male/Female	All surveys	All surveys
Family form	Now we'd like to ask you about who you live with. Not everyone lives with both their parents. Sometimes people live with just one parent, sometimes they have two homes or two families.	Yes/No: Please fill in column A for your main or your only home. Fill in column B if you have a second home (not including holiday or summer houses). Mother; Father; Stepmother (or father's girlfriend); Stepfather (or mother's boyfriend); Grandmother; Grandfather; I live in a foster home or children's home; Someone or somewhere else: please write it down (open-ended).	All surveys	HBSC
	Please say how many brothers and sisters live here (including half, step or foster brothers and sisters). Please write in the number or write 0 (zero) if there are none.	Scale: Both homes are included Brothers and Sisters.	All surveys	HBSC
Migrant status	Were your parents born in this country?	Yes/No	All surveys (except HBSC)	EQLS, PIRLS, TIMSS
	Where you born in this country?	Yes/No	All surveys (except HBSC)	EQLS, PIRLS, TIMSS

Source: Adapted from survey items listed in Annex 5.

350. The recommendations here are naturally limited to what is considered to be feasibly possible for all survey coordinators to achieve. It is an example of the minimum harmonisation needed to make fuller use of the datasets available to researchers and policymakers. It is accepted that any final list would need to be agreed between survey providers, and it is further accepted that it does not, on its own, meet recommendation by the EC Social Protection Committee and others for an effective monitoring tools for child well-being (not least because it does not include series data that would otherwise be collected as part of these efforts – see SPC, 2012).

351. Of all of the recommendations in Table 26, the recommendations for questions on family form may be the most difficult to agree because of the time taken to complete this type of item (as well as the space it takes up in the survey). HBSC surveys collects detailed information on the family structure of both first and second homes, and also asks children are not presently living in a children's home. Although this means sacrificing time and space in other surveys, it would seem to be very relevant considering the trend in single and reformed families is growing, and social policies designed to adapt for children living in more than one household are becoming more common.

5.5.3.2 Harmonising socio-economic measures

352. Following efforts to come up with a minimal set of socio-demographic indicators to harmonise surveys, an effort could be made to match socio-economic items of material deprivation and parental education and employment (see Table 27).

353. Child deprivation items are asked of in most of the surveys (see Annex 5), normally in lists which are later composited for reporting purposes, and so a limited set of agreed items could be found and added without much effort to the surveys lacking them (this is also a lot less complicated than asking PISA to undertake questions on risk, or morbidity, and HBSC be unlikely to ever undertake an literacy assessment type module). The added value of harmonising deprivation items would be opportunity it would provide to expand substantially social gradient work (a key recommendation of this report and of the SPC report [2012]).

354. Parental employment and education data is also ever present in the surveys reviewed, and if it were successfully matched with child data, its inclusion may facilitate harmonisation of child well-being data across child and household surveys.

Table 27: Suggested socio-demographic items for harmonisation across surveys

Matching indicator	Subject or question	Item detail	Survey use similar item	Survey use exact item
Deprivation	About how many books are there in your home?	Tick box (0-10; 11 to 25; 26-100; 101-200; 200 or more).	PISA, TIMSS	PIRLS
	Do you have a personal computer with internet connection?	Yes/No	EU-SILC, HBSC, PIRLS, PISA, TIMSS	None
	Does your family own a car, van or truck?	Ordinal List (No, Yes one, more than one).	EU-SILC, HBSC, PIRLS, PISA, TIMSS	HBSC
	Some young people go to school or to bed hungry because there is not enough food at home. How often does this happen to you?	Ordinal List (Always, Often, Sometimes, Never).	HBSC	HBSC
	Do you have your own bedroom for yourself?	Yes/No	EU-SILC, HBSC, PIRLS, PISA, TIMSS	HBSC
	Do you have enough equipment at home to do all homework tasks?	Yes/No	EU-SILC, HBSC, PIRLS, PISA, TIMSS	None
	During the past 12 months, how many times did you travel away on holiday (vacation) with your family?	Ordinal List (Not at all, Once, Twice, More than twice).	EU-SILC, HBSC, PIRLS, PISA, TIMSS	HBSC
Parental education	Father's level of education?	Tick box: ISCED 1-6	All surveys (except PIRLS)	
	Mother's level of education?	Tick box: ISCED 1-6	All surveys (except PIRLS)	
Parental Employment	What does your father (main male guardian) do in his main job? (examples including Unemployment, My parent is sick, not applicable)	Open ended	All surveys (except PIRLS)	
	What does your mother (main female guardian) do in her main job? (examples including Unemployment, My parent is sick, not applicable)	Open ended	All surveys (except PIRLS)	

Source: Adapted from survey items listed in Annex 5.

5.6 Which indicators can be used, now, for monitoring

355. What information can be used for monitoring the well-being of children in developed countries today, if none of the recommendations above are actioned?

356. As part of their on-going work on child poverty and child well-being, the Social Protection Committee of the European Commission have identified the following areas as possible priorities for further data and indicators' development (SPC, 2012: 68-69):

- Children in particularly vulnerable situation (in particular migrant children, children outside traditional types of households, e.g. in alternative care).
- Child deprivation (an area in which work is on-going at time of writing).
- Children's health (in particular in relation to health inequalities, social gradients)
- Access, affordability and quality of childcare and early childhood education services (on-going in the context of EU SILC revision)
- Child's social participation (including recreational, cultural, civic and sporting activities) has been identified as an area that is important to children's rights, development and future outcomes.
- Additional poverty measures, including food and fuel poverty.

357. In terms of **vulnerable children**, and as has been noted many times in this document, the SPC observes that surveys designed to capture child well-being do not have great success in capturing the most vulnerable because these children are not in school, or part of standard household respondent groups. The SPC report highlights migrant children, disabled children, and children in-out-of-home settings as vulnerable groups. There are data available on these groups, although besides migrant status (see Annex 5 table on Socio-demographics for many items on migrant status), these are either non-mandatory and so do not cover sufficient countries, or asked only of adults (see examples in ESS, EQLS and EU-SILC for disability/chronic illness in Annex 5's Health table). The example of HBSC's question to children about living in an out-of-home setting (see Table 28) is unique, and could be added to other school surveys. It is clearly evident that much more could be done in the area of capturing information on the most vulnerable children in cross-national survey work, if ambitions to robustly and regular monitor the well-being of vulnerable children are to be met.

358. How to build a child deprivation indicator is an ongoing priority for Net-SILC2 research, and the EUROSTAT taskforce on Material Deprivation, as stated in the SPC report (SPC 2012: 54). Regarding **child deprivation measures**, the review in chapter 4 (see section 4.4.3) highlights the relative wealth of information in this area. The recommendation to use this topic to harmonise surveys also shows the importance of such measures for explaining the variation seen in child well-being outcomes monitored in on-going surveys. Moreover, despite the minimalist approach to measuring child deprivation outlined in Tables 27 and 28, and cautions regarding its insufficiency, it is beyond this report to recommend a full and comprehensive set of child deprivation indicators, and pre-empt the recommendation of the EUROSTAT taskforce. Rather it is possible to stress that child deprivation monitoring is possible, and it should be age-related – reflecting on developmental and gender-based needs where relevant – and attempts to draw up a final list from available data in future work should be distinct from household deprivation measures and involve input from children themselves whenever possible.

359. In relation to **children's health and inequalities**, or health by social gradient, and acknowledging the many series that could be used but not discussed in this survey report (the SPC report lists infant mortality, as well as health service indicators), ESPAD and HBSC can be used to identify health indicators in the child population in Europe (form age 11 through to 16). For outcomes by social gradient, harmonisation with surveys with better parental income, work and education indicators would improve the quality of social gradient tests undertaken with these sources. The indicators used in the UNICEF Innocenti Report Card 10 *The children left behind* (UNICEF, 2010) report card are listed in Table 28, under health inequalities, although many other health (objective and subjective self-reported measures) are available in the surveys. For associated social gradient measures see Table 27, and depending on source or potential for harmonisation, Annex 5 (the Socio-demographic table for parental education, and the Income and deprivation table for other social gradient measures).

Table 28: Suggested items for complementing Social Protection Committee monitoring recommendations

Indicator	sub-indicator	Survey use	Subject or question	Item detail
Vulnerable Children	Family form	HBSC	Now we'd like to ask you about who you live with. Not everyone lives with both their parents. Sometimes people live with just one parent, sometimes they have two homes or two families.	Yes/No: I live in a foster home or children's home.
Child Deprivation	Deprivation (cultural)	e.g. PIRLS, TIMSS, PISA	About how many books are there in your home?	Tick box (0-10; 11 to 25; 26-100; 101-200; 200 or more).
	Deprivation (IT)	e.g. EU/SILC, TIMSS, PISA	Do you have a personal computer / internet connection?	Yes/No
	Deprivation (household)	e.g. EU-SILC, HBSC	Does your family own a car, van or truck?	Ordinal List (No, Yes one, more than one).
	Deprivation (food)	e.g. HBSC	Some young people go to school or to bed hungry because there is not enough food at home. How often does this happen to you?	Ordinal List (Always, Often, Sometimes, Never).
	Deprivation (personal space)	e.g. HBSC	Do you have your own bedroom for yourself?	Yes/No
	Deprivation (educational)	e.g. PISA	Do you have enough equipment at home to do all homework tasks?	Yes/No
	Deprivation (family hol./leisure)	e.g. HBSC	During the past 12 months, how many times did you travel away on holiday (vacation) with your family?	Ordinal List (Not at all, Once, Twice, More than twice).
Children's Health	Health complaints	HBSC	In the last 6 months: how often have you had the following...?	Likert scale (About every day, More than once a week, About every week, About every month, Rarely or never): Headache; Stomach-ache; Back ache; Feeling low; Irritability or bad temper; Feeling nervous ; Difficulties in getting to sleep; Feeling dizzy.
	Healthy eating	HBSC	How many times a week do you usually eat or drink…?	Likert scale (Never, < once a week, once a week, 2-4 days a week, 5-6 days a week, once a day, every day, and Every day, more than once): Fruits; Vegetables.
	Vigorous physical activity	HBSC	Over a typical or usual week, on how many days are you physically active for a total of at least 60 minutes per day?	Likert scale: Also asks for recent week data.
Childcare and Early Childhood Education Services	attendance	PISA	Did you attend pre-school <ISCED 0>	Yes/No:
	costs	EU-SILC	Number of hours of educaiton during a usual week (pre-school)/ number of hours of child care during a usual week.	Reported in hours (0-99).
Child social participation	Participation (political opinion)	ICCS	How interested are you in political issues?	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
	Participation (social issues)	ICCS	How interested are you in social issues?	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
	Participation (work/civic)	ESS	In the last month have you done any paid/voluntary work?	Tick box
	Participation (personal social)	HBSC	How many evenings per week do you usually spend out with your friends?	Scale (None through until 7).
	Participation (IT/social)	HBSC	How often do you talk to your friend(s) on the phone or send them text or email messages?	Likert scale (Rarely or never, 1 or 2 days a week, 3 or 4 days a week, 5 or 6 days a week, Every day).
	Participation (sports / activities)	PIRLS / TIMSS	How often do you? Play sports	Likert scale
Additional poverty measures	Food poverty	See deprivation (food) above.		
	Debt / Arrears	EQLS	Has your household been in arrears in the last 12 months on? Rent or mortgage / utility bills	Yes/No
		EU-SILC	Arrears on loans?	Yes/No
		EU-SILC	Arrears on mortgage payment?	Yes/No

Source: see Annex 5.

360. For the available information on **childcare access, affordability and quality**, it is only EU-SILC that provides information on childcare use, and it is only PISA that refers to attendance of childcare in their questionnaires (or of ISCED 0). Using this information, inference can be drawn about quality by relating childcare attendance in PISA to literacy outcomes, or in relation to access, and its impact on employment in households with children, using EU-SILC statistics. Questions that assess the links between childcare

access and quality to child development are most effectively addressed in longitudinal studies. Questions regarding affordability are better addressed via efforts to record payments, benefits and fees policies in series derived from administrative data, or micro-simulation modelling.

361. **Child social participation** information is more readily available than might otherwise be assumed, but is mostly restricted to surveys with lower country coverage than EU-SILC or PISA, and in the case of ICCS (so far no second wave is planned) may not be useable for trend monitoring purposes. Table 28 lists participation in the forms of: political and social opinion, work and civic engagement (though this is from ESS meaning a youth subsample is likely to suffer from very small numbers), participation in social activities both personal and virtual, as well as sporting activities. Important forms of children's participation are missing from this review, despite this information being collected in the past. CIVED in 1999 collected information on whether children were actively participating in political and social groups (the ICCS data reported here only covers opinions), and is an example for work developing more participation items in future surveys (see Annex 5 – Civic participation and time use). PISA, in 2000, asked their respondents about the extent of their cultural participation (attending theatre, music concerts, sporting events and so on), but has not since repeated this item in compulsory questionnaires. In 2009, the module of child deprivation in EU-SILC included items on children's regularity of activities and participation in school trips, but this one-off module has yet to be repeated (SPC, 2012).

362. **Alternative poverty measures** mentioned in the SPC report (2012) refer to food and fuel poverty. Table 28 picks up on this suggestion and lists present attempts to measure food poverty in the HBSC study (also covered in deprivation), and information about arrears and debt (which includes reference to utility bills, and so risk to supplies of necessities to the home). The arrears and debt figures complement well efforts to measure deprivation and income poverty as they are not included in wealth or income measures (including poverty rates calculations). Debt and arrears can create situations in family homes that have an impact on living conditions (including stress and relationship breakdown) above those identified by income poverty alone, as they express how well families can meet basic needs with the means available to them.

363. It is important to note, at this stage, that these indicators can be used to meet the recommendation of monitoring priorities set by the SPC, but they do not represent a comprehensive set of indicators for monitoring. For instance they do not include children's education, parental employment or education, income statistics etc., and for the most part cover only children in the school surveys – the majority of whom are older children. A comprehensive set of monitoring indicators would address these shortfalls (partly addressed in the SPC's monitoring framework) and draw from other sources such as series data. Recommendations from this report on additional data needs, such as information on children's time use, relationships and issues of neglect and protection are dealt with in section 5.3.

5.6.1 Issues of coverage and timeliness in available monitoring data

364. For each of the proposals above there are issues of country coverage and timeliness of collection and reporting to consider. Country coverage varies by surveys, for only PISA covers all OECD countries, and only EU SILC covers all EU member states (participation by country is covered in chapter 2). The indicators derived from child surveys are only available every three or four years, and generally with a two year lag. Household survey-derived data are annual in the case of EU-SILC, but again a lag of up to 18 months can be expected. But despite the inevitable gaps and limitations with the available data, it is necessary to begin monitoring and attempt to improve the coverage (concepts and countries) and timeliness or reporting during the process of monitoring, in order to avoid waiting for 'the perfect time' to begin work, and the potential this would bring for missing important messages and lessons that could inform policy and practise in the meantime.

365. Key steps to improve the coverage of child well-being indicators, and country coverage, are outlined in other parts of this recommendations chapter, and cover issues such as harmonisation and the development of new surveys. Importantly, and worth reiterating, is the need to use the data available at the moment, whilst continually striving to improve the knowledge base on child well-being and living conditions of all children.

5.7 Summary

366. Despite good evidence in support of the robustness of much child well-being data (from both children's surveys and household surveys), methods to test for validity and the robustness of such data before using it in cross-national comparisons, and methods to adjust the data for non-response and other limitations, there remains a number of important limitations that stifle the achievement of the 'gold standard' in evidence based policy for all children. Missing populations in the data (for example children out of school or children under 9), the lack of a single harmonised database covering the main dimensions of child well-being, restrictions to access, and the lengthy turnaround of much of this data, all point to the need for new data collection initiatives in the field of child well-being.

367. Although, there remain important gaps in the knowledge of children's well-being cross-nationally, this project has achieved a number of important steps. First, the review in Chapter 4 has provided an empirical grounding for recommending the adaptations to content in existing surveys. Second, Chapters 3 and 4 have provided empirical grounding useful for understanding how amendments to the position, or wording of times can affect comparability, as well as a reflection of the statistical form in which responses are gathered and collated, for possible applications in research (particularly distributional analysis), and the first steps in efforts to understand and recommend methods for harmonisation in the field.

368. Whilst researchers and policymakers are waiting for new comparative data it is important that the best is made from data which is available. And in doing so, a careful balance has to be made between the need to inform policies for children today, and inform policies for children accurately. Nonetheless, the need for new data, by age and by topic, is so very obvious that investment in new collections, and investment at the margins of present surveys, is necessary if real steps forward in this area are to be made. If policymakers, survey coordinators (present and future), and researchers can communicate joint goals, roles, needs and limitations better – and make some of the steps recommended in this final chapter – the future for effective monitoring of comparative child well-being, for the benefit of all children, can be secured.

REFERENCES

Blum, A., Goldstein, H., and Guérin-Pace, F. (2001) International Adult Literacy Survey (IALS): an analysis of international comparisons of adult literacy, Assessment in Education: Principles, Policy & Practice, 8: 2, 225 - 246.

Bradshaw, J., and Richardson, D. (2009) An Index of Child Well-being in Europe. Child Indicators Research, Vol. 2, No. 3, pp 319 - 351.

Bradshaw, J., Hoelscher, P., & Richardson, D. (2007). An Index of Child Well-being in the European Union. Social Indicators Research, 80(1), 133-77.

Currie, C., et al. (eds) (2008) Inequalities in Young People's Health in Context. HBSC International report from the 2001/2002 study, WHO Regional Office for Europe. Copenhagen.

Currie, C., Roberts, C., Morgan, A., Smith, R., Settertobulte, W., Samdal, O. and Barnekow Rasmussen, V. (eds) (2004) 'Young People's Health in Context. Health Behaviour in School-Aged Children Study (HBSC)': International report from the 2001/2002 study, WHO Regional Office for Europe.

D'Orazio M., Di Zio M., Scanu M. (2011) *Statistical Matching and Imputation of Survey Data with the Package StatMatch for the R Environment*. ISTAT, Italy.

EQLS (2010), "European quality of Life Survey (EQLS)". (www.eurofound.europa.eu/areas/qualityoflife/eqls/), December 2010.

ESPAD (2010) ESPAD - European School Project on Alcohol and other Drugs website. Accessed May 2010 at http://www.espad.org/

ESS (2010), "European Social Survey (ESS)". (www.europeansocialsurvey.org), December 2010.

EU SILC (2010), "The European Survey on Income and Living Conditions (EU SILC)". (epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc), December 2010.

European Agency for Development in Special Needs Education (2008 and 2010), SNE Country Data. www.european-agency.org/

European Commission (2008), Social Protection Committee (2008). Child poverty and well-being in the EU: Current status and way forward, Luxembourg: Office for Official Publications of the European Communities. http://www.libertysecurity.org/IMG/pdf_ke3008251_en.pdf. Ganzeboom, H., et al (1992) 'A standard international socioeconomic index of occupational status.' Social Science Research Vol. 21, pp 1-56.

Förster M. and Richardson D. (2011) "Réduction de la pauvreté des enfants : comparaisons internationales", Politiques sociales et familiales n° 104, Cnaf, juin 2011.

Freeman, R. B., S. Machin, and M. Viarengo (2010) Variation in Educational Outcomes and Policies Across Countries and of Schools within Countries. NBER Working Paper 16293. Massachusetts, USA

Hanuschek, E. & Woessmann, L. (2010) Sample Selectivity and the Validity of International Student Achievement tests in Economic Research. IZA Discussion Paper No. 4926, Germany.

HBSC (2010) HBSC - Health Behaviour in School-aged Children Study website. Accessed May 2010 at http://www.hbsc.org/

ICCS (2010) ICCS - International Civic and Citizenship Education Study website. Accessed May 2010 at http://iccs.acer.edu.au/

Micklewright, J., Schnepf, S., and Skinner, C. (2010) Non-response Biases in Surveys of School Children: The Case of the English PISA Samples. IZA Discussion Paper No. 4789, Germany.

Mullis, I. V. S et al (2009) PIRLS 2011 Assessment Framework. TIMSS & PIRLS International Study Centre, Boston.

OECD (2009), Doing Better for Children, OECD, Paris (www.oecd.org/els/social/childwellbeing).

OECD (2011) Doing Better for Families, OECD, Paris.

OECD PISA (2011) OECD PISA Survey data (What PISA produces?), OECD, Paris available at: http://www.oecd.org/pages/0,3417,en_32252351_32236130_1_1_1_1_1,00.html.

OECD/PISA (2010) PISA - Programme for International Student Assessment website. Accessed May 2010 at http://www.oecd.org/pages/0,3417,en_32252351_32235731_1_1_1_1_1,00.html

PIRLS (2010) PIRLS - Progress in International Reading Literacy Study website. Accessed May 2010 at http://timss.bc.edu/

Rose and Özcan, (2007), First European Quality of Life Survey: Quality of Life in Turkey. EQLS, Dublin.

SPC - Social Protection Committee of the European Commission (2012)SPC Advisory Report to the European Commission on Tackling and Preventing Child Poverty, Promoting Child Well-being. Social Protection Committee, Brussels, 27th of June, 2012.

Steiner, P. and Cook, D. (undated) Matching and Propensity Scores, Chapter forthcoming in *The Oxford Handbook of Quantitative Methods*.

Stiglitz, J. E., A. Sen, and J. Fitoussi (2010) Report by the Commission on the Measurement of Economic Performance and Social Progress, Accessed November 2010, www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

Tarki/Applica (2010) Child Poverty and Child Well-being in the European Union: Report for the European Commission. Tarki Social Research Institute, Budapest.

The Children's Society (2012) The Good Childhood Report 2012: A review of our children's well-being, http://www.childrenssociety.org.uk/sites/default/files/tcs/good_childhood_report_2012_final.pdf

TIMSS (2010) TIMSS - Trends in International Mathematics Science Study website. Accessed May 2010 at http://timss.bc.edu/

UNCRC (1990) United Nations Convention on the Rights of the Child, United Nations, New York. www.unhchr.ch/html/menu3/b/k2crc.htm.

UNESCO (2011) Institute for Statistics: Education data, available at: http://www.uis.unesco.org/Pages/default.aspx

UNICEF IRC (2007) Child poverty in perspective: An overview of child well-being in rich countries, Innocenti report card 7, UNICEF: Florence.

UNICEF IRC (2010) The Children Left Behind: A league table of inequality in child well-being in the world's richest countries, Innocenti report card 9, UNICEF: Florence.

ANNEX 1: COORDINATORS INVOLVED

369. A number of survey coordinators have been contacted to comment on the results of the project. Comments received have been incorporated into the final version of the document. These include:

ESPAD – European School Project on Alcohol and other Drugs
Contact: Dr Thor Bjarnason
Faculty of Social Science and Law
University of Akureyri
Iceland

HBSC – Health Behaviour in School-aged Children Study
Contact: Prof. Candace Currie
CAHRU (Child and Adolescent Health Research Unit)
University of St Andrews, Scotland

PIRLS – Progress in International Reading Literacy Study
Contact: Dr Michael Martin and Dr Ina Mullis
Co Directors of the TIMSS and PIRLS International Study Centre
Boston College, MA, USA.

PISA – Programme for International Student Assessment
Contact: OECD PISA
Education Directorate
2 Rue Andre Pascal
Paris – France.

TIMSS – Trends in International Mathematics Science Study
Contact: Dr Michael Martin and Dr Ina Mullis
Co Directors of the TIMSS and PIRLS International Study Centre
Boston College, MA, USA.

ANNEX 2: CHILD SURVEYS

ESPAD – European School Survey Project on Alcohol and Other Drugs.

Origin of the study	Systematic information on alcohol and drug use prevalence rates is usually collected with epidemiological surveys. In several countries large-scale series of school surveys on alcohol and other drugs use are being conducted among students. Despite this, it might be difficult to do cross-national comparisons due to differences in methodology, especially regarding the data collection instrument (the questionnaire). For this reason, a school survey expert group was formed in the mid-1980s within the Pompidou Group at the Council of Europe. Their task was to develop a standard questionnaire that could be used by countries that wanted to perform national school surveys resulting in comparable data. In the early 1990s, Björn Hibell and Barbro Andersson at the Swedish Council for Information on Alcohol and Other Drugs (CAN) initiated a collaborative project by contacting a number of European researchers to explore the possibilities of doing simultaneously performed school surveys with a common methodology. The European School Survey Project on Alcohol and Other Drugs, the ESPAD project was planned to build on the experiences of the Pompidou school survey group. The Pompidou Group made funds available to support the participation of researchers from the former Eastern Europe countries in a first Project meeting. In March 1994 representatives from 21 countries met at the Council of Europe, Strasbourg, for the first project meeting. Methodological aspects as well as the questionnaire were thoroughly discussed. Questions and details remaining unsolved after the meeting were handled by a working group. ESPAD data collections are performed every four years. The first data collection was carried out in 1995, the following in 1999, 2003, 2007 and the latest in 2011. The sixth data collection will be carried out in 2015.
Management Structure	
Statement of purpose.	Due to the lack of comparable data on substance abuse among European teenagers, the ESPAD was initiated to collect comparable data on alcohol, tobacco and drug use among 15-16 year old students in as many European countries as possible. The most important goal in the long run is to monitor time trends in alcohol and drug habits among students in Europe and to compare trends between countries and between groups of countries over time (Hibell & Andersson, 2006).
Funded by	The Swedish Ministry of Health and Social affairs, the Swedish National Institute of Public Health, the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) and the Pompidou Group at the Council of Europe have supported the ESPAD-project throughout the years. In addition, each country nominates a Principal Investigator, who oversees the project at the national level and is expected to raise funds for national data collection. Some other organisations, such as the Swedish International Development Cooperation Agency (SIDA), the Social Research Unit for Alcohol Studies in Finland (THL, former STAKES) and the Alcohol Research Council of the Swedish Alcohol Retailing Monopoly (SRA) have also provided funds at some point of the project.
Affiliations.	CAN in Stockholm has been the coordinating institution of the ESPAD-project since the project was initiated. Each participating country holds the responsibility for the project as a whole as well as at the national level. Every participating country appoints a Principal Investigator who oversees and holds the responsibility for the project at the national level.
Consortium	
Government partners	Swedish Ministry of Health and Social Affairs
Expert panels	Developments and amendments to the ESPAD questionnaire are undertaken by the questionnaire group. **In 2007, the questionnaire group** consisted of: Chair, Björn Hibell, Sweden; Thoroddur Bjarnason, Iceland; Ludwig Kraus, Germany; Patrick Lambrecht, Belgium; Leena Metso, Finland; and Alfred Uhl, Austria. **Thematic groups** In September 2010, there were 7 active thematic groups. Each group consists of approximately seven people. The themes for these groups are Antisocial behaviour, Former Soviet Union countries, Polydrug use, Risk group, Risky Drinking Habits, Substance Use in Small Islands, and Multilevel Analysis. Terminated Thematic groups include Attitudes, Gender, and Trends. For information about the members of these groups, see: *http://www.espad.org/Uploads/Documents/Overview_of_the_ESPAD_Project.pdf*
Data set basic information.	

Dates.	1995, 1999, 2003, 2007 and 2011
Countries.	In 2003 35 Countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Faroe Islands, Finland, France, Germany, Greece, Greenland, Hungary, Iceland, Ireland, Isle of Man, Italy, Latvia, Lithuania, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, Turkey, Ukraine, United Kingdom. In 2007 35 Countries: Armenia, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Faroe Islands, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Isle of Man, Italy, Latvia, Lithuania, Malta, Moldavia, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Slovakia, Slovenia, Sweden, Switzerland, Ukraine, United Kingdom. In 2011 37 countries: new countries in the latest survey were Albania and Liechtenstein. In addition, three countries (Bosnia and Herzegovina, Kosovo and the Netherlands) conducted the survey but for different reasons.
Contents.	The questionnaire contains core as well as optional questions. All countries should employ the core questions, which include: socio-demographic background variables, alcohol, tobacco and drug related questions as well as questions for methodological checking purposes. Data is also collected on leisure time use, educational attainment, and missing school. The 2011 questionnaire contains four modules: “Integration”, “Psychosocial”, “Cannabis use” and “Deviance” Countries are welcome to include one or more modules in the questionnaire as well as a small number of country specific questions of special interest. Field-testing the questionnaire is highly recommended for countries joining the ESPAD project. A translation-back translation process is also highly recommended.
Core data includes:	The standard questionnaire contains core questions on use of alcohol, tobacco and drugs.
Contextual data includes:	The standard questionnaire contains socio-demographic background variables.
Requirements of access.	The use of new database is limited to ESPAD researchers for the first two years. Since 2003, datasets from all participating countries have been merged and stored into common databases, Access to ESPAD 1995, 1999, 2003 and 2007 data is freely available online at www.espad.org.
Methodology	
Sources and collection methods.	Data is collected through the administration of a standardised school questionnaire.
Unit of analysis.	Students that turn 16 years old during the calendar year of the data collection. Students who are unable to understand or for other reasons cannot answer the questionnaire without assistance - e.g. children with special educational needs or severely handicapped - are not included in the target population.
What is the sample design?	The goal of the sampling process is to obtain a national and gender wise representative data set and it is up to each national ESPAD researcher to find the optimal method of sampling. The number of participating students is however suggested not to be below 2.400, which allows breakdowns by sex and another variable. Sampling should consist of randomly selected classes.
Sample threshold?	In 2011, the average response rate was 87%, only Malta and Romania had rates slightly below 80%
Collection window	March/April is the recommended period for data collection.
Delivery in the field	
Training and management of field staff	The Swedish Council for Information on Alcohol and Other Drugs (CAN) is the co-ordinator of the ESPAD-project and publishes the international reports. All major decisions are taken at the annual Project meetings, which are open to all ESPAD researchers. Principal Investigators and Contact Persons have the right to vote in these meetings. They include decisions about the questionnaire, the data collection procedure, Regional seminars, national project plans and reports, as well as about the ESPAD data base. Between the Project meetings all major decisions are taken by the Steering Committee (the Working group), elected at Project Meetings. Besides taking necessary decisions the Steering Committee also prepares the Project meetings. The committee also functions as an editorial board for the ESPAD reports, which means that while the reports mainly have been written by the coordinators, manuscripts have been sent to the committee members for comments and amendments. It has been important for the ESPAD project that the Steering Committee should include experienced researchers from different parts of Europe. In 2012, the **Steering Committee** consists of: • Chair, Björn Hibell, Sweden;

	• Ulf Guttormsson, Sweden; • Salme Ahlström, Finland; • Olga Balakireva, Ukraine; • Thoroddur Bjarnason, Iceland; • Anna Kokkevi, Greece; and • Ludwig Kraus, Germany. The daily coordination of ESPAD is done by the two Coordinators**:** Björn Hibell and Ulf Guttormsson from the Swedish Council for Information on Alcohol and Other Drugs (CAN), Sweden. Besides the coordination they prepare the Steering Committee and Project meetings**.**
Translation and other procedures	A translation-back translation process is also highly recommended.
Length of test	
Replacement sampling	
What modules are included*?*	The questionnaire contains core as well as optional question (See appendix IV for ESPAD 2011 questionnaire: http://www.espad.org/Uploads/ESPAD_reports/2011/The_2011_ESPAD_Report_FULL_2012_10_29.pdf).
Can countries add their own items?	Countries are welcome to include one or two modules in the questionnaire as well as a small number of country specific questions of special interest.
Next wave/project.	
Next date and availability.	Next data collection will be carried out in 2015.
New content / questions.	
Countries to be added.	In ESPAD 2007 two new countries, Armenia and Moldova. In ESPAD 2011, another two countries: Albania and Liechtenstein
Previous waves/projects.	ESPAD 1995, 1999, 2003, 2007 and 2011
Previous content not repeated in the next wave.	There were only a few modifications in the 2011 questionnaire compared to the previous one and the vast majority of these modifications were in the optional segment of the questionnaire. The questionnaire used in the 2007 survey differed to some extent from the template used for the three first waves. There were some structural changes and changes of wording of some questions.
Countries not repeating the study in the next wave.	In ESPAD 2011, only five German states participated in the survey compared to seven in 2007. Data for the Russian Federation included only Moscow as opposed to the whole country in 2007. In addition, in the Isle of Man data was collected but results could not be delivered.
Links	
Home page for the website	http://www.espad.org
To access the data online	A selection of key variables (selected questions) of ESPAD 1995, 1999, 2003 and 2007 is available online at http://www.espad.org/en/Keyresult-Generator/
To access the international report	The ESPAD 1995, 1999, 2003, 2007 and 2011 reports are available online at . http://www.espad.org/en/Reports--Documents/
How to access the technical reports	
To access national reports	*http://www.espad.org/en/References--Literature/*
To access a list of research that uses the data	ESPAD data available online at http://www.espad.org/en/Data-Collection/Databases/ Research based on ESPAD data available online at http://www.espad.org/en/References--Literature/
Contact email:	bjorn.hibell@can.se

HBSC – Health Behaviour in School-aged Children

Origin of the study	The Health Behaviour in School-aged Children (HBSC) study was initiated in 1982 by researchers from three countries and shortly afterwards the project was adopted by the World Health Organization as a WHO collaborative study. The Child and Adolescent Health Research Unit (CAHRU) of the University of St. Andrews, Scotland is currently the International Coordinating Centre (ICC) of HBSC. The HBSC study is a cross-national research project conducted by an international network of research teams in collaboration with the WHO Regional Office for Europe. Its aim is to gain new insight into and to increase understanding of young people's health, well-being, health behaviour and social context. Researchers from three countries started the HBSC study in 1982; since then, a growing number of countries and regions have joined the study network (WHO 2004).
Management Structure	
Statement of purpose.	To chart the links between the life circumstances and young people's health and build a better understanding of the factors that influence their well-being, for WHO's Member States and the European Region as a whole. The study seeks new insight into adolescents' health, health behaviour and lifestyles in the social context. It examines young people aged 11, 13 and 15 years in order to inform and influence health promotion and health education policy and practice for young people at the national and international levels, and to advance scientific knowledge. Though time-series analysis is not an explicit goal, HBSC has completed three identical waves that enable evaluation of trends over time.
Funded by	For the HBSC 2009/10, data collection in each country or region was funded at the national level. This and other financial support was offered by the various government ministries, research foundations and other funding bodies in the participating countries and regions. WHO supports research dissemination, publication of international reports, etc. HBSC country teams also pay an annual subscription that contributes to international coordination and data management. The ICC is funded by NHS Health Scotland, plus country subscriptions.
Affiliations.	WHO Regional Office for Europe (for dissemination of the international reports), the Child and Adolescent Health Research Unit (CAHRU) of the University of St Andrews, and the University of Bergen, Norway
Co-ordinators	The Child and Adolescent Health Research Unit (CAHRU) of the University of St Andrews
Government partners	Some governments fund national teams, NHS Health Scotland funds ICC.
Expert panels	The HBSC 2009/10 report has several authors comprising the HBCS Principle Researchers and Team members, with input from the members of the editorial group of the HBSC international research network. The authors of the chapters collaborated in the development of the HBSC protocol, working together in topic-related focus groups, whose approaches depended on the authors' different disciplinary backgrounds and familiarity with related research paradigms. HBSC editorial board include experts from • Child and Adolescent Health Research Unit, University of St Andrews, Scotland, United Kingdom; • School of Medicine, University of St. Andrews; • HBSC Scientific Development Group; • WHO Regional Office for Europe; • University of Hertfordshire, Hatfield, United Kingdom; • Faculty of Social and Behavioural Sciences, Utrecht University, Netherlands; • Knowledge and Analytical Services, Welsh Government, United Kingdom (Wales); • Department of Health Promotion and Development, University of Bergen, Norway; The expert panels include the following: • Co-ordinating committee • Scientific Development Group • Methodology development Group • Policy Development Group • Focus Groups of experts on particular topic areas • Principal Investigator Assembly
Data set basic information.	

Dates.	1983/84, 1985/1986, 1989/90, 1993/94, 1997/1998, 2001/02, 2005/06 and 2009/10
Countries.	At present, membership of HBSC is restricted to countries and states within the WHO European region (and the USA and Canada). HBSC 2000/01, 36 countries: Finland, Norway, Austria, Belgium (French), Hungary, Israel, Scotland, Spain, Sweden, Switzerland, Wales, Denmark, Canada, Latvia, Poland, Belgium (Flemish), Czech Republic, Estonia, France, Germany, Greenland, Lithuania, Russia, Slovak Republic, England, Greece, Portugal, Rep. of Ireland, USA, Macedonia, Netherlands, Italy, Croatia, Malta, Slovenia, Ukraine. HBSC 2005/06, 41 countries: The same countries adding Bulgaria, Iceland, Luxembourg, Romania and Turkey. HBSC 2009/10: The same countries as in 2005/6 with the addition of Albania and Armenia
Contents.	HBSC is a standard school-based survey with data collected through self-completion questionnaires administered in the classroom. The HBSC survey instrument is a standard questionnaire developed by the international research network and used by all participating countries. Each survey questionnaire contains a core set of questions looking at the following: • *Background factors*: demographics and maturation, social background (family structure, socio-economic status); • *Individual and social resources*: body image, family support, peers, school environment; • *Health behaviours*: physical activity, eating and dieting, smoking, alcohol use, cannabis use, sexual behaviour, violence and bullying, injuries; • *Health outcomes*: symptoms, life satisfaction, self-reported health, Body Mass Index There are a series of standard optional packages that expand on particular focus areas which can be included in national questionnaires Many countries also include additional items in their national questionnaire that are of particular interest on a national level
Core data includes:	Core data includes individual and social resources, health behaviours and health outcomes (see contents).
Contextual data includes:	Contextual data includes socio-demographic background.
Requirements of access.	Access to HBSC data is not freely available. The international data file is restricted for the use of member country teams for a period of three years from its completion. After this time the data is available for external use by agreement with the International Coordinator and the Principal Investigators. The latest accessible data is from the 2005/06 survey for researchers, and 2009/10 for organisations by prior agreement. Access is given where proposed work does not overlap with ongoing HBSC research activities and plans. Access can be refused if the proposed work is not collaborative in nature.
Methodology	
Sources and collection methods.	Data is collected through the administration of a standard student questionnaire. The school-based survey collects data through self-completion questionnaires, administered in the classroom. In some countries an optional school level questionnaire is used. This comprises a set of core items on the school demographics.
Unit of analysis.	Students aged 11, 13 and 15 years old.
What is the sample design?	The sampling strategy used in each country aimed to achieve nationally representative groups at the mean ages of 11.5, 13.5 and 15.5. The sample consists of approximately 1500 from each age group (i.e. a total of 4500 from each participating country). For the 2005/06 survey children were selected using a clustered sampling design, where the initial sampling unit was either the school class or the school. The latter was sampled when class lists were not available. The requirement for minimum recommended sample size was met in the majority of countries and regions. The 2009/10 wave includes stratification and schools are sampled based on "probability-proportional-to-size" (PPS). Some countries have also collected the information via a census
Sample threshold?	*1500 respondents per age group or census*
Collection window	For HBSC 2005/06, fieldwork for each cross-national survey is carried out over a period of around seven to eight months, from October to May. However, more precisely, within each country the actual time of survey was only around 2 complete months; fieldwork time is determined by school entry dates and so differs by country
Delivery in the field	

Training and management of field staff	The Child and Adolescent Health Research Unit (CAHRU) of the University of St Andrews is currently the International Coordinating Centre (ICC) of HBSC. The ICC is responsible for coordinating all international activities within the HBSC research network. These include the production of survey protocols and international reports; planning and organising the network's bi-annual meetings; facilitating network communications; and acting as a resource centre for information on the study for members and external agencies and professionals. Training and management of field staff is carried out in each country according to the international protocol. Materials for use in schools are provided in the international protocol -see 2009/10 protocol for further details. The data collected in each country is sent to the HBSC Data Management Centre at the University of Bergen, Norway. It is then cleaned and compiled into an international data file by the Norwegian Social Science Data Services (NSD).
Translation and other procedures	See above. Notes….. Forward and back translation from English by different translators. Specialist team receive back translations and discuss discrepancies with national teams. The desired aim is for meaning to be the same This is explained better in the new Protocol -I have included the text in the section 2 edits
Length of tests	
Replacement sampling	
What modules are included?	The survey consists of one standard survey (see contents above). There are also optional packages focusing in more detail on particular areas. In many cases these optional packages provide scales
Can countries add their own items?	Yes, many countries also include additional items in their national questionnaire that are of particular interest on a national level.
Next wave/project.	
Next date and availability.	HBSC survey 2009/10 is currently available for analysis.
New content / questions in 2009/10?	2013/14
Countries to be added.	Additional countries in HBSC 2005/06 compared to HBSC 2000/01 are Bulgaria, Iceland, Luxembourg, Romania and Turkey.
Previous content (from 2000/01) not repeated in the next wave.	neighbourhood + others possibly –I can get this detail if you want
Countries not repeating the study in the next wave.	Bulgaria was unable to get funding for 2009/10 survey but remains a member of the network
Links	
Home page for the website	http://www.hbsc.org/
To access the data online	Data not freely available online.
To access the international report	Aggregate data is available in the 1993/94, 1997/98, 2000/2001, 2005/6 and 2009/10 international report which is available online at http://www.hbsc.org/publications/international/
How to access the technical reports?	*By request*
To access national reports	Aggregate data for individual countries is available is available on the country websites. For access to the countries website see http://www.hbsc.org/countries.html.
To access a list of research that uses the data	Research based on HBSC data available online through http://www.hbsc.org/publications.html
Contact email:	cahru@st-andrews.ac.uk or info@hbsc.org

ICCS – International Civic and Citizenship Education Study

Origin of the study	The Civic Education Study (CIVED) was a civics and citizenship study undertaken by IEA in 1999. CIVED assessed how prepared children were for their roles as citizens in modern society through constructing a meaningful and comparable international test of students' citizenship knowledge. The study also recorded attitudes and civic behaviours in participating countries. Ten years later the IEA International Civic and Citizenship Education Study (ICCS) will build on CIVED, assessing not only the changes in civic knowledge that have occurred in the past ten years, but investigating the recent and new challenges for educating children for their roles as citizens. ICCS is designed to compare cross-nationally how prepared young people are to undertake their roles as citizens. The study will report student achievement on a test of conceptual understandings and competencies in civic and citizenship education at the eighth grade (children should not be younger however than 13.5 years).
Management Structure	
Statement of purpose.	The purpose of the International Civic and Citizenship Education Study (ICCS) is to investigate, in a range of countries, the ways in which young people are prepared and consequently ready and able to undertake their roles as citizens. The study will report student's conceptual understandings and competencies in civics and citizenship. The survey will collect data on activities and attitudes related to civic and citizenship education from students. Student, teacher, school and national questionnaires will provide contextual data related to family, school and country for each respondent.
Funded by	Participating countries and the IEA currently fund the ICCS. For the development of the project IEA received support from the UNESCO Education Sector, Division for the Promotion of Quality Education, Section of Education for Peace and Human Rights. The international participation fee is USD 30,000 per year for the four-year duration of the study, or USD 120,000 in total. Countries participating in the regional modules receive financial support from the European Commission (European Module) and the Inter-American Development Bank (Latin American Module) (http://www.iea.nl/ICCS.html).
Affiliations	CRELL (Institution of the European Commission) for European countries and CERLALC for Latin American countries
Co-ordinators	A consortium of three organisations co-ordinate the ICCS 2009: the Australian Council for Educational Research (ACER), the National Foundation for Educational Research (NFER – UK based) and the Laboratorio di Pedagogia sperimentale (LPS) at the Roma Tre University (Italy). A joint management committee made up of consortium partners, IEA Secretariat and IEA DPC oversee the overall management of ICCS 2009. The yearly meeting of the National Research Coordinators reviews project on the subsequent stages.
Government partners	Ministries of education of the participating countries.
Expert panels	The project advisory committee consists of experts from within the consortium, CRELL, CERLALC and independent citizenship experts. The ICCS 2009 Project Advisory Committee includes members from the following institutes: • Australian Council for Educational Research, Australia • IEA Secretariat • IEA Technical Executive Group • University of London, United Kingdom • Ministry of Education, Chile • University of Leiden, the Netherlands • Centre for Research on Lifelong Learning, European Commission • CERLALC • University of Maryland, United States • Hong Kong Institute of Education • University of Liege, Belgium

	ICCS also makes use of Expert Consultants at appropriate points in the study, selected on the basis of expertise, relevant to specific aspects of need.
Data set basic information	
Dates.	2008/2009
Countries.	Austria, Belgium (Flemish), Bulgaria, Chile, Chinese Taipei, Colombia, Cyprus, Czech Republic, Denmark, Dominican Republic, England, Estonia, Finland, Greece, Guatemala, Hong Kong SAR, Indonesia, Ireland, Italy, Korea, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Mexico, Netherlands, New Zealand, Norway, Paraguay, Poland, Russian Federation, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Thailand.
Contents.	The contents of the ICCS 2009 are categorised into two parts of an Assessment Framework. The first part contributes the core data in the form of a civic and citizenship framework. The civic and citizenship framework collects the outcome data measured through a cognitive test and a student's perceptions questionnaire. The second part contributes the contextual framework, providing for the collection of contextual factor, such as socio-demographics of students and their families and classroom environments, including teaching and learning to help explain the variation in the outcome measures.
Core data includes:	Student civic and citizenship knowledge (outcome) data is collected using a 45 minute International cognitive test and a 25 minute International student perceptions questionnaire (a region-specific cognitive test of around 15 minutes and a region-specific questionnaire of around 15 minutes is administered in countries participating in the European and Latin American regional modules; in countries participating in the Asian regional module a questionnaire of up to 30 minutes is administered). Four content domains make up the civic and citizenship content to be assessed by cognitive test and questionnaire items: Civic Society and systems – covering aspects of the roles, rights and responsibilities of citizens, and the State and Civil institutions involved. Civic Principles – covering aspects of the shared ethical foundations of societies, in particular sub-domains of Equity, freedom and social cohesion. Civic Participation – assessing understanding of individuals actions in their communities including decision-making, influencing, and community participation. Civic Identities – covering the civic roles of individuals and perceptions of these roles. Civic self-image and civic connectedness are sub-domains here.
Contextual data includes:	Contextual data is collected from students using an International student background questionnaire, which takes around 15 minutes to complete. Student contextual data includes items covering student characteristics (like sex and age), family background (parental socio-economic status, country of birth, language), school climate (student-teacher relations, classroom practices, student influence at school), and civic-related activities both in school and the wider community. Teachers, schools and National Research Coordinators also provide contextual data. A teacher questionnaire of up to 30 minutes covers teachers' perceptions of conditions for civic and citizenship education at their school. Teacher characteristics and reports on their participation are among other variables to be collected. The questionnaire is designed to be administered to all teachers, and not just teachers of citizenship education. A school questionnaire, administered to principals, covers school-level variables such as level of autonomy and school characteristics (enrolment, teacher numbers and so on). A questionnaire will also be distributed (on-line) to National Research Coordinators and covers national level data on the background for the teaching of civic and citizenship education in each participating country.
Requirements of access.	Access to ICCS data will be available for free at http://rms.iea-dpc.org/#. The planned release date for the ICCS database is November 2010. For more details on the content of the ICCS 2009 please see the publication list on the ICCS website http://iccs.acer.edu.au/index.php?page=publications-and-papers .
Methodology	

Sources and collection methods.	Data are collected through the administration of test and questionnaires to students and questionnaire to their teachers and school principals. In addition, an on-line questionnaire for national research centres collects data on the national context for civic and citizenship education.
Unit of analysis	Eighth grade students in participating countries are principal units of analysis. Units of analysis also include schools and education systems. Target populations are children in their 8th year of schooling provided that their average age is above 13.5 years of age. Educational systems that took part in CIVED in 1999 and assessed students in a different grade have been asked to include this grade as well in order to enable comparisons.
What is the sample design?	The typical sample size for each country will be about 150 schools and between 3,500 to 5,000 students. Probability Proportionate by Size (PSS) sampling is used to select the schools. Within schools, typically one classroom is randomly sampled to take the tests. Exclusions are made in the case of three types of student: 1) Intellectually disabled students 2) functionally disabled students not able to perform in ICCS testing situation 3) non-native language speakers with less than one year instruction in the test language
Sample threshold	85% of response school participation rates (before replacement), as well as student-level and teacher-level participation rates of 85%. Overall participation rates should be no lower than 75%.
Collection window?	The ICCS 2009 was initiated in October 2006. The Field Trial took place in October to December 2007 and Main Survey was planned for October - December 2008 (Southern Hemisphere) and February - May 2009 (Northern Hemisphere).
Delivery in the field	
Training and management of field staff	The Australian Centre for Education Research (ACER) serves as the International Study Centre for the IEA Civic and Citizenship Education Study. Staff at the International Study Centre are responsible for the design and implementation of the international test, student questionnaires, analyses and reporting. In addition, ACER also coordinated the development of the Latin American and Asian regional module instruments. The National Foundation for Educational Research (NFER) in the UK has the main responsibility for the European regional module instrument and the National Context Survey. Teacher and school questionnaires for the international test are the main responsibility of Laboratorio di Pedagogia sperimentale (LPS) at the Roma Tre University, Italy. Sampling activities, field procedures and data management are coordinated by the IEA Data Processing and Research Center (DPC) in Hamburg Germany. Quality monitoring and translation verification procedures and relationships with participating countries will be managed via the IEA secretariat in Amsterdam. Further in each country, a team made up of a national representative, called the National Research Coordinator (NRC), data entry and management staff and translation specialists are responsible for implementing ICCS in accordance with international procedures.
Translation and other procedures	See above
Length of tests	See above
What modules are included?	There are three regional modules with region-specific student instruments: The European regional module (25 countries), the Latin American regional module (six countries) and the Asian regional module (five countries). In addition, there are shorter optional sections in the student and teacher questionnaires.
Can countries add their own items?	Countries may add a limited amount of questionnaire material after international and regional (if applicable) test and questionnaire items.
Next wave/project	
Next date and availability.	
New content / questions.	
Countries to be added.	

Previous content not repeated in the next wave.	
Countries not repeating the study in the next wave.	
Links	
Home page for the website	http://iccs.acer.edu.au/
To access the data online	Data is available to the public via the IEA Study data repository at the following website: http://rms.iea-dpc.org/#.
To access the international report	The International report, and regional reports are available as pdf documents through the ICCS website (http://iccs.acer.edu.au/).
How to access the technical reports	The technical report is available as pdf document at the ICCS website (http://iccs.acer.edu.au/).
To access national reports	See the ICCS website (http://iccs.acer.edu.au/).
To access a list of research that uses the data	See the ICCS website (http://iccs.acer.edu.au/).
Contact email:	iccs@acer.edu.au

PIRLS – Progress in International Reading Literacy Study.

Origin of the study	Beginning with the 1970/71 International Study of Reading Comprehension in 15 countries, and continuing with the 1991 International Reading Literacy Study in 32 countries, the International Association for the Evaluation of Educational Achievement (IEA) has pioneered international comparative studies of reading achievement. At the beginning of the new century, IEA re-focused its research program in reading literacy with the establishment of the Progress in International Reading Literacy Study (PIRLS), a regular assessment of 4th grade reading achievement on a 5-year cycle. Conducted in 35 countries, PIRLS 2001 was the first cycle of the PIRLS program, providing a wealth of information about reading achievement and the home, school, and classroom environment for the teaching and learning of reading. PIRLS 2006 was the second study in the PIRLS cycle, collecting data on 4th grade reading achievement and the context for learning reading in 40 countries (PIRLS 2006 Encyclopaedia, 2007). PIRLS was expanded in 2011 to include prePIRLS, a less difficult evaluation that was implemented in select countries.
Management Structure	
Statement of purpose.	To monitor trends in reading literacy of children in middle childhood. PIRLS provides comparative information on the reading literacy of students in their fourth year of formal schooling, with a particular focus on the factors, at home and in the school, which facilitates the acquisition of literacy in young children (PIRLS Assessment Framework and Specifications 2006).
Funded by	The National Centre for Education Statistics (NCES) of the US Department of Education, the World Bank, Boston College, the National Foundation for Educational Research in England and Wales, and fees paid by the participating countries (PIRLS 2011 Encyclopedia and PIRLS 2011 Assessment Framework)
Affiliations.	IEA Secretariat in Amsterdam, Statistics Canada in Ottawa, the IEA Data Processing Center in Hamburg, and Educational Testing Service in Princeton, New Jersey. In each country, a national representative, called the National Research Coordinator (NRC). (PIRLS Assessment Framework 2011).
Co-ordinators	PIRLS is coordinated by the TIMMS and PIRLS International Study Center (ISC) located at Boston College, United States and the International Association for the Evaluation of Educational Achievement (IEA), Amsterdam, the Netherlands. The PIRLS executive directors are Michael Martin and Ina V.S. Mullis, and the IEA executive director is Hans Wagemaker.
Government partners?	US Department of Education
Expert panels	Updating the PIRLS Framework and Specifications for the 2006 assessment was a collaborative effort involving a series of reviews by the Reading Development Group, the Questionnaire Development Group, and the National Research Coordinators. **The Reading Development Group** • Isa al-Nasheet, Ministry of Education, Bahrain ;Jan Mejding, Danish University of Education, Denmark; • Sue Horner, Qualifications and Curriculum Authority, England; • Elinor Saiegh-Haddad, Bar-Ilan University, Israel; • Pierre Reding, Ministere de l'Education Nationale, Luxembourg ; • Galina Zuckerman, Russian Academy of Education, Russian Federation; • Elizabeth Pang, Ministry of Education, Singapore; • Caroline Liberg, Uppsala University, Sweden; • Elois Scott, United States; and • Mohamed Al-Mekhlafy, United Arab Emirates University, Yemen. **The Questionnaire Development Group** • Hwawei Ko, National Central University, Chinese Taipei;Marc Colmant, Ministère de l'Éducation Nationale, France; • Knut Schwippert, University of Hamburg, Germany; • Megan Chamberlain, Ministry of Education, New Zealand; • Ragnar Gees Solheim, National Centre for Reading, Education and Reading Research, University of Stavanger, Norway; • Abdessalem Bouslama, Supreme Education Council, Qatar; • Sarah Howie, University of Pretoria, South Africa; and • Valena Plisko, National Center for Education Statistics, United States. The **National Research Coordinators** are national representatives who are responsible for implementing PIRLS in accordance with international procedures.
Data set basic information.	
Dates.	2001, 2006, and 2011.

Countries.	PIRLS 2001, 35 countries: Argentina, Belize, Bulgaria, Canada (Ontario, Quebec), Colombia, Cyprus, Czech Republic, England, France, Germany, Greece, Hong Kong, Hungary, Iceland, Iran, Israel, Italy, Kuwait, Latvia, Lithuania, Macedonia, Moldova, Morocco, Netherlands, New Zealand, Norway, Romania, Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, Sweden, Turkey, United States. PIRLS 2006, 40 countries: Austria, Belgium (Fl.), Belgium (Fr.), Bulgaria, Canada, Chinese Taipei, Denmark, England, France, Georgia, Germany, Hong Kong, Hungary, Iceland, Indonesia, Iran, Israel, Italy, Kuwait, Latvia, Lithuania, Luxembourg, Macedonia, Moldova, Morocco, Netherlands, New Zealand, Norway, Poland, Qatar, Romania, Russian Federation, Scotland, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Trinidad and Tobago, United States. PIRLS 2011, 48 countries: Australia, Austria, Azerbaijan, Belgium (Fr.), Botswana, Bulgaria, Canada, Chinese Taipei, Colombia, Croatia, Czech Republic, Denmark, England, Finland, France, Georgia, Germany, Honduras, Hong Kong, Hungary, Indonesia, Iran, Ireland, Italy, Kuwait, Lithuania, Malta, Morocco, Netherlands, New Zealand, Northern Ireland, Norway, Oman, Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Trinidad and Tobago, United Arab Emirates, United States. PrePIRLS, 3 countries: Botswana, Colombia, South Africa.
Contents.	Contents of PIRLS 2011 include assessments of reading literacy, which encompasses reading for literary experience, reading to acquire and use information, and four types of comprehension processes. The survey also includes background questionnaires. Following a matrix sample design, the reading assessments focus on reading purposes and comprehension processes, consisted of ten passages (five literary and five informational passages). Students' understanding of these texts was assessed through multiple choice and constructed responses. The background questionnaires consisted of a student questionnaire (home and school experiences including basic demographic information, attitudes towards reading, computer use and literacy resources in the home, school climate for learning), a home survey (language spoken in the home, preschool experiences, homework activities, home-school involvement, books in the home, parents' education and involvement, early literacy and numeracy activities, parents' reading activities, and attitudes towards reading), a teachers questionnaire (their background and education, class characteristics, the school climate for learning, attitudes towards teaching, student engagement, homework, and school resources), and a school questionnaire (including school and local area characteristics, instructional time, parental involvement, teaching staff, reading education characteristics and school resources and climate).
Core data includes:	The survey includes reading literacy (curricula measure), focussing on reading purposes (literacy experience, and acquire and use of information) and comprehension processes (focus on and retrieve explicitly stated information; make straight forward inference; interpret, and integrate ideas and information; examine and evaluate, content, language, and textual elements).
Contextual data includes:	PIRLS administers background questionnaires to students, their parents, their teachers, and the principals of their schools. The questions are designed to measure key aspects of students' home and school environments (see contents above).
Requirements of access.	Access to PIRLS data 2001 available for free at http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_UG.html Access to PIRLS data 2006 available for free at http://timss.bc.edu/pirls2006/ Access to PIRLS data 2011 is forthcoming online.
Methodology	
Sources and collection methods.	Data was collected through the administration of literacy assessment questionnaires and background questionnaires in school to pupil, teachers and administrators and at home to parents.
Unit of analysis.	Students, primary care givers of students (parents), teachers and schools.
What is the sample design?	A two-stage stratified clustering sample was used. Schools were first sampled randomly proportionate to size. (Probability proportionate to size increases the likelihood of larger schools being selected in the sample. One of the following measures of size was used: the average student enrolment in the fourth grade, the number of classrooms in the fourth grade, or the total student enrolment in the school). Students are then sampled by classes (cluster sampling) in school which are sampled probability proportionate to size. Students are randomly selected to take parts of the assessment using booklets designed for matrix sampling.

Sample threshold?	85 percent of both the schools and students, or a combined rate (the product of schools' and students' participation) of 75 percent.
Collection window	For PIRLS 2006, the tests were administered between October-December 2005 for PIRLS 2006 data collection in Southern Hemisphere countries. March-June 2006 PIRLS 2006 data collection was conducted in Northern Hemisphere countries.
Delivery in the field	
Training and management of the field staff	The TIMSS & PIRLS International Study Center at Boston College serves as the International Study Center for IEA's studies in mathematics, science, and reading – the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). The staff at the Study Center is responsible for the design and implementation of the study. In each country, a national representative, called the National Research Coordinator (NRC), is responsible for implementing PIRLS in accordance with international procedures.
Translation and other procedures	The Questionnaire Development Group finalizes the questionnaires one year to 9 months before the completion of data collection (following field tests).
Lengths of tests	Ten 40-minute blocks of passages, with breaks between sections, for a total of six hours.
Replacement sampling	
What modules are included?	Reading for literary experience and reading to acquire and use information modules. PIRLS retained six passages and items from the 2001 and 2006 assessments – three literary and three informational – to be included in the PIRLS 2011 assessment. As in 2006, four new passages were selected.
Can countries add their own items?	No.
Next wave/project.	
Next date and availability.	Survey in 2015-2016, with data released in 2017.
New content / questions.	N/A
Countries to be added.	N/A.
Previous content not repeated in the next wave.	N/A
Countries not repeating the study in the next wave.	N/A
Previous waves/projects.	The IEA's Reading Literacy Study 1991 is the closest comparator to the 2001, 2006, and 2011 waves. The PIRLS reading test and student questionnaires from 1991 were re-administered in nine countries. Together the surveys were used to study trends in reading literacy in nine countries over ten years. Countries involved that administered identical research projects for each year (1991 and 2001) were Greece, Hungary, Iceland, Italy, New Zealand, Singapore, Slovenia, Sweden, and the United States.
Links	
Home page for the website	http://www.pirls.org/
To access the data online	IEA Online Data Analyzer: www.iea.nl/iea_studies_datasets The international database for PIRLS 2006 data will be available online in February 2008 at http://www.pirls.org/ , under "International Databases"
To access the international report	Aggregate country data is available in the PIRLS 2001 report online at http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_IR.html. Aggregate country data is available in the PIRLS 2006 report online at
How to access the technical reports	The PIRLS 2001 and 2006 technical reports are available online at www.pirls.org
To access national reports	

To access a list of research that uses the data	Research based on PIRLS data is available online at www.pirls.org, under “Publications.”
Contact email:	isc@bc.edu

PISA – Programme for International Student Assessment.

Origin of the study	Responding to member countries' demands for regular and reliable data on the knowledge and skills of their students and the performance of their education systems, the OECD began work on the Programme for international Assessment (PISA) in the mid-1990s. PISA was officially launched in 1997, with the first survey taking place in 2000, the second in 2003 and the third in 2006. Future surveys are planned in 2009, 2012.
Management Structure	
Statement of purpose.	PISA aims to measure how far students approaching the end of compulsory education have acquired some of the knowledge and skills essential for full participation in the knowledge society. In addition to monitoring the outcomes of education systems in terms of student achievement in the three main subject areas of reading, mathematics and science, within an international framework, PISA will seek to deepen policy insights by: • Developing better ways of tracking student progress, including exploring the possibility of comparing student progress between primary education and the age of 15 years; • Allowing closer comparisons between performance and instruction, by extending the option of examining student performance within a single grade and gathering more data about classroom experiences; • Making use of computer-based assessments, not only to measure Information and Communication Technology (ICT) literacy skills but also to allow for a wider range of dynamic and interactive tasks and to explore more efficient ways of carrying out the main tests of student knowledge and skills in reading, mathematics and science. These innovations will initially be explored by countries as a supplementary, optional component of PISA, but where appropriate may be integrated into the survey's core.
Funded by	OECD member countries
Affiliations.	
Consortium	The design and implementation of the present surveys, within the framework established by the Pisa Governing Board (PgB), is the responsibility of an international consortium led by the Australian Council for Educational research (ACER). Other partners in this consortium include the National Institute for Educational Measurement (CITO) in the Netherlands, WESTAT and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NRIES) in Japan.
Government partners	
Expert panels	PISA represents a collaborative effort among the OECD member governments to provide a new kind of assessment of student achievement on a recurring basis. The assessments are developed co-operatively, agreed by participating countries, and implemented by national organisations. The **PISA governing Board** (PgB), representing all nations at the senior policy levels, determines the policy priorities for PISA in the context of OECD objectives and oversees adherence to these priorities during the implementation of the programme. Participating countries take responsibility for the project at the policy level. Experts from participating countries also serve on working groups, the **Science Expert Group, Reading Expert Group and Mathematics Expert Group**, that are charged with linking the PISA policy objectives with the best available substantive and technical expertise in the field of internationally comparative assessment. **Science Expert Group** • Chair, Rodger Bybee, Biological Sciences Curriculum Study Colorado Springs, United States; • Ewa Bartnik, University of Warsaw, Warsaw; • Poland Peter, Fenshammonash University, Queensland, Australia; • Paulina Korsnakova, Department of educational measurement, Bratislava, Slovak republic • Robert laurie, Department of Education of New Brunswick, New Brunswick, Canada; • Svein Lie, University of Oslo Blindern, Norway; • Pierre Malléus, Ministry for National Education, Higher education and Research Champigneulles, France; • Michelina Mayer National Institute for the Evaluation of Instructional Systems, Rome, Italy; • Robin Millar University of York, York, United Kingdom; • Yasushi Ogura, National Institute for Educational Policy Research Tokyo, Japan, • manfred Prenzel University of Kiel, Kiel, Germany; and

	• Andrée Tiberghien, University of Lyon, Ste Foy les Lyon, France. **Reading Expert Group** • Chair, John de Jong Language Testing Services, Oranjestraat, Netherlands; • Irwin Kirsch, Educational Testing Service, Princeton, New Jersey, United States; • Marilyn Binkley, National Center for Educational Statistics, Washington, d. c., United States; • Alan Davies University of Edinburgh, Scotland, United Kingdom; • Stan Jones, Statistics Canada, Nova Scotia, Canada; • Dominique Lafontaine, University of Liège, Liège, Belgium; • Pirjo Linnakylä, University of Jyväskylä, Jyväskylä, Finland; • Martine Rémond IUFM de Créteil, University of Paris 8, Andresy, France. **Mathematics Expert Group** • Chair Jan de Lange Utrecht University Utrecht, Netherlands; • Werner Blum, University of Kassel, Kassel, Germany; • John Dossey, Illinois State, University Eureka, Illinois, United States; • Zbigniew Marciniak, Warsaw University, Warsaw, Poland; • Mogens Niss IMFUFA, Roskilde University, Roskilde, Denmark; and • Yoshinori Shimizu, University of Tsukuba, Tsukuba-shi, Ibaraki, Japan.
Data set basic information.	
Dates.	2000, 2003 and 2006
Countries.	PISA 2003: 41 countries participated in the assessment: Australia, Austria, Belgium, Brazil*, Canada ,Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong-China*, Hungary, Iceland, Indonesia*, Ireland, Italy, Japan, Korea, Latvia*, Liechtenstein, Luxembourg, Macao-China*, Mexico, The Netherlands, New Zealand, Norway, Poland, Portugal, Russian Federation*, Serbia and Montenegro*, Slovak Republic, Spain, Sweden, Switzerland, Thailand*, Tunisia*, Turkey, United Kingdom, United States, Uruguay*. PISA 2006: 57 countries participated in the assessment: Argentina*, Australia, Austria, Azerbaijan*, Belgium, Brazil*, Bulgaria*, Canada, Chile*, Colombia*, Croatia*, Czech Republic, Denmark, Estonia*, Finland, France, Germany, Greece, Hong Kong-China*, Hungary, Iceland, Indonesia*, Ireland, Israel*, Italy, Japan, Jordan*, Korea, Kyrgyz Republic*, Latvia*, Liechtenstein*, Lithuania*, Luxembourg, Macao-China*, Mexico, The Netherlands, New Zealand, Norway, Poland, Portugal, Qatar*, Republic of Montenegro*, Republic of Serbia*, Romania*, Russian Federation*, Slovak Republic, Slovenia*, Spain, Sweden, Switzerland, Chinese Taipei*, Thailand*, Tunisia*, Turkey, United Kingdom, United States, Uruguay*. The above countries are Members of the OECD, except those marked with an asterisk (*).
Contents.	PISA covers the domains of *reading, mathematical* and *scientific literacy* not so much in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life. Emphasis is placed on the mastery of processes, the understanding of concepts and the ability to function in various situations within each domain. PISA assesses student performance beyond the confines of reading, mathematics and science, as cross curricular competencies like ICT skills, communication and problem-solving will be important to students' futures. In each three-yearly PISA survey, one subject (domain) is chosen as a focus while two other subject areas have been assessed more briefly. This allowed, for each subject area, a detailed profile of what a country's students can do every nine years, and an update of their performance every three years. For PISA 2006 a total of about 390 minutes of test items is covered, with different students taking different combinations of test items. Pencil-and-paper tests are used, with assessments lasting a total of two hours for each student. Test items are a mixture of multiple-choice items and questions requiring students to construct their own responses. The items are organized in groups based on a passage setting out a real-life situation. A total of about seven hours of test items is covered, with different students taking different combinations of test items. PISA administers background questionnaires to students (approximately 20-30 minutes). The questionnaires seek information about: Students and their family backgrounds, including their economic, social and cultural capital; Aspects of students' lives, such as their attitudes towards learning, their habits and life inside school, and their family environment; Aspects of schools, such as the quality of the schools' human and material resources, public and private control and funding, decision-making processes, and staffing practices; context of instruction, including institutional structures and types, class size, and the level of parental involvement; Strategies of self-regulated learning, motivational preferences and goal orientations, self-related cognition mechanisms, action control strategies, preferences for different types of learning situations, learning styles, and social skills required for co-

	operative or competitive learning; Aspects of learning and instruction in science, including students' motivation, engagement and confidence with science, and the impact of learning strategies on achievement related to the teaching and learning of science. Further school principals are given a 20-minute questionnaire about their schools. Two additional questionnaires are offered as international options: • A computer familiarity questionnaire focusing on: *i)* availability and use of information and communications technology (ICT), including the location where Ic is mostly used as well as the type of use; *ii)* ICT confidence and attitudes, including self-efficacy and attitudes towards computers; and *iii)* learning background of ICT, focusing on where students learned to use computers and the Internet. • A parent questionnaire focusing on a number of topics including the student's past science activities, parents' views on the student's school, parents' views on science in the student's intended career and the need for scientific knowledge and skills in the job market, parents' views on science and the environment, the cost of education services, and parents' education and occupation.
Core data includes:	Core data is gathered through a questionnaire on each of the three domains of reading, mathematical and scientific literacy, whereby one is chosen as a focus. The background questionnaires are an integrated part of the core data as well.
Contextual data includes:	PISA administers background questionnaires to students and the principals of their schools. The questions are designed to measure key aspects of students' home and school environments.
Requirements of access.	Access is free and available online at http://www.pisa.oecd.org/
Methodology	
Sources and collection methods.	Data was collected through the administration of background questionnaires and assessment questionnaires. Pencil-and-paper tests are used, with assessments lasting a total of two hours for each student. Not all students answer all questions in the assessment. The PISA 2006 test units are arranged in 13 clusters, with each cluster designed to occupy 30 minutes of testing time. There are seven science clusters, two reading clusters and four mathematics clusters. The clusters are placed in 13 booklets, according to a rotated test design. Each booklet contains four clusters and each student is assigned one of these two-hour booklets. There is at least one science cluster in each booklet (Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006: 13).
Unit of analysis.	Students (the international target population was consequently defined as all students aged from 15 years and 3 (completed) months to 16 years and 2 (completed) months at the beginning of the assessment period) and school principals.
What is the sample design?	The sample design for PISA is referred to as a *two-stage stratified sample*. The first-stage sampling units consist of individual schools having 15-year-old students. Schools are sampled systematically with probabilities proportional to a measure of size (PPS), with the measure of size being a function of the estimated number of eligible (15-year-old) students enrolled. The comprehensive national list of all eligible schools is called the school sampling frame. Prior to sampling, schools in the sampling frame can be assigned to a predetermined number of either implicitly or explicitly defined strata. A minimum of 150 schools will be selected in each country, with the requirements of national options often requiring a somewhat larger sample. As the schools are sampled, replacement schools are simultaneously identified, should they be needed to replace non-participating sampled schools. The second-stage sampling units are students within sampled schools. Once schools are selected to be in the sample, a list of each sampled school's 15-year-old students must be prepared. From each list, 35 students will be selected with equal probability (all 15-year-old students will be selected if fewer than 35 are enrolled). Note that 30 students (150 schools x 30 students = 4500 students) has been inflated to 35 to account for an anticipated combined no response and exclusion rate of 15% among students within schools. The number of students selected per school can be varied from the figure of 35, as a national option. However, if fewer than 35 students per school are to be selected, then: a) the sample size of schools must be increased beyond 150, so as to ensure that at least 4,500 students are assessed, and b) the number of students selected per school must be at least 20, so as to ensure adequate accuracy in estimating variance components within and between schools (an analytical objective of PISA) (School Sampling Preparation Manual PISA 2003 Main Study, Version One, June 2002: 7-8).
Sample threshold	Tests are typically administered to between 5000 and 10,000 students from at least 150 schools in each country.
Collection window?	For PISA 2000 the international requirement was that the assessment had to be conducted during a 42-day period

	between 1 March 2000 and 31 October 2000 For PISA 2003, 2006 and 2009.
Delivery in the field	
Training and management of field staff	PISA represents a collaborative effort among the OECD member governments to provide a new kind of assessment of student achievement on a recurring basis. The assessments are developed co-operatively, agreed by participating countries, and implemented by national organisations. The PISA governing Board (PgB), representing all nations at the senior policy levels, determines the policy priorities for PISA in the context of OECD objectives and oversees adherence to these priorities during the implementation of the programme. Participating countries take responsibility for the project at the policy level. Experts from participating countries also serve on working groups (Science Expert Group, Reading Expert Group and Mathematics Expert Group) that are charged with linking the PISA policy objectives with the best available substantive and technical expertise in the field of internationally comparative assessment. The design and implementation of the present surveys, within the framework established by the PgB, is the responsibility of an international consortium led by the Australian council for educational research (ACER). Other partners in this consortium include the National Institute for Educational Measurement (CITO) in the Netherlands, WESTAT and the Educational Testing Service (ETS) in the United States, and the National Institute for Educational Policy Research (NRIES) in Japan. The study is implemented in each country by a National Project Manager (NPM) who implemented the procedures prepared by the Consortium. To implement the assessment in schools the NPMs were assisted by School Coordinators and Test Administrators. Each NPM typically had several assistants, working from a base location that is referred to throughout this report as a 'national centre'.
Translation and other procedures	
Length of test	
Replacement sampling	
What modules are included?	
Can countries add their own items?	Yes, two additional questionnaires are offered as international options (see Contents).
Next wave/project.	
Next date and availability.	For PISA 2009, 62 countries have already signed up.
New content / questions.	PISA 2009 will include Educational Career questionnaire and a parental questionnaire.
Countries to be added.	PISA 2006 included 17 more countries compared to PISA 2003: Azerbaijan, Bulgaria, Chile, Colombia, Croatia, Estonia, Israel, Jordan, Lithuania, Qatar, Republic of Montenegro, Republic of Serbia, Romania, Chinese Taipei.
Previous content not repeated in the next wave.	
Countries not repeating the study in the next wave.	
Previous waves/projects.	PISA 2000 and PISA 2003.
Previous content / questions and countries.	The main focus of PISA 2000 was on reading literacy, it included an extensive set of tasks in this domain. In PISA 2003, the emphasis was on mathematical literacy and an additional domain on problem solving was introduced. PISA 2006 cycle focuses on scientific literacy.
Links	
Home page for the website	http://www.pisa.oecd.org/
To access the data online	The international database for PISA 2000 data is available online at http://pisaweb.acer.edu.au/oecd/oecd_pisa_data.html The international database for PISA 2003 data is available online at http://pisaweb.acer.edu.au/oecd_2003/oecd_pisa_data.html

	The international database for PISA 2006 data is available online at http://pisa2006.acer.edu.au/
To access the international report	http://www.oecd.org/document/2/0,3343,en_32252351_32236191_39718850_1_1_1_1,00.html
How to access the technical reports	The PISA 2000 technical report is available online at http://www.pisa.oecd.org/document/7/0,3343,en_32252351_32236159_33688711_1_1_1_1,00.html The PISA 2003 technical report is available online at http://www.pisa.oecd.org/document/13/0,3343,en_32252351_32236173_35188685_1_1_1_1,00.html The PISA 2006 technical report is available online at http://www.oecd.org/document/41/0,3343,en_32252351_32236191_42025897_1_1_1_1,00.html
To access national reports	Aggregate country data is available in the participating PISA 2000 country reports available online at http://www.pisa.oecd.org/document/3/0,3343,en_32252351_32236159_33680899_1_1_1_1,00.html Aggregate country data is available in the participating PISA 2003 country reports available online at http://www.pisa.oecd.org/document/62/0,3343,en_32252351_32236159_34575550_1_1_1_1,00.html Aggregate country data is available in the participating PISA 2006 country reports available online at http://www.pisa.oecd.org/document/....
To access a list of research that uses the data	Research based on PISA 2000 data available online at http://www.pisa.oecd.org/document/61/0,3343,en_32252351_32236159_33680573_1_1_1_1,00.html Research based on PISA 2003 data available online at http://www.pisa.oecd.org/document/10/0,3343,en_32252351_32236173_34342538_1_1_1_1,00.html
Contact email:	edu.pisa@oecd.org

TIMSS – Progress in International Reading Literacy Study.

Origin of the study	The International Association for the Evaluation of Educational Achievement (IEA) has, over the past 50 years, conducted comparative research studies focusing on educational policies, practices, and outcomes in more than 60 countries around the world. Organized around a Secretariat located in Amsterdam, The Netherlands, and a data processing center in Hamburg, Germany, IEA, through its various projects, continues to report on a wide range of topics and subject matters. Since 1995 the IEA's Trends in International Mathematics and Science Study (TIMSS) collects internationally comparable data of student achievement in mathematics and science on a regular four-year cycle.
Management Structure	
Statement of purpose.	TIMSS is designed to help countries all over the world improve student learning in mathematics and science. It collects educational achievement data at the fourth and eighth grades to provide information about trends in performance over time together with extensive background information to address concerns about the quantity, quality, and content of instruction. It must be noted that TIMSS studies the effectiveness of curriculum and instruction in relation to student achievement across times and across grades. TIMSS therefore is designed to provide important information for policy development, to foster public accountability, to allow areas of progress or decline in achievement to be identified and monitored, and to address concerns for equity.
Funded by	The international management of the TIMSS 2011 project is funded by the participating countries with support from the World Bank and the National Center for Education Statistics at the United States Department of Education. Boston College and the U.K.'s National Foundation for Educational Research provide support. For countries participating at both fourth and eighth grades, the participation fee is USD 60,000 per year for the four-year duration of the study, or USD 240,000 in total. The fee for participating at the fourth or eighth grade only is USD 40,000 per annum, for a total of USD 160,000 (http://www.iea.nl/timss2007.html).
Affiliations.	*The World Bank*
Co-ordinators	As previous TIMSS assessments, TIMSS 2007 is coordinated by the IEA International Study Center (ISC) located at Boston College, United States. The study Co-directors are Dr Michael Martin and Dr Ina Mullis, and the IEA Executive Director is Hans Wagemaker. Other members of the TIMSS 2007 consortium are the IEA Secretariat and the IEA Data Processing Center, Statistics Canada, and the Educational Testing Service.
Government partners	*United States Department of Education*
Expert panels	The staff at the TIMSS & PIRLS ISC is responsible for the design and implementation of these studies. In developing and producing the TIMSS frameworks, ISC staff conducted a collaborative effort involving a series of iterative reviews by the TIMSS 2007 Science and Mathematics Item Review Committee and the National Research Coordinators of participating countries. The Science and Mathematics Item Review Committee (SMIRC) worked with staff from the International Study Center in developing all aspects of the frameworks and particularly the mathematics and science frameworks. They made recommendations for the content and cognitive domains, and focus areas for policy-orientated research. **Mathematics experts** • Kiril Bankov, University of Sofia, Bulgaria; Karen Manriquez, Ministry of Education, Chile;Fou-Lai Lin, National Taiwan Normal University, Chinese Taipei; • Khattab Mohammad Ahmad AbuLibdeh, National Center for Human Resources Development, Jordan; • Robert Garden, New Zealand; Christoph Selter, TU Dortmund Mathematics Department, Germany;Liv Sissel Grønmo, University of Oslo, ILS Norway; • Mary Lindquist, Columbus State University, United States; • Hung-Hsi Wu, University of California, Berkeley, United States **Science experts** • Martina Kekule, Charles University in Prague, Czech Republic; • Jouni Viiri, University of Jyvaskyla, Finland; Saule Vingeliene, Educational Development Centre, Lithuania; Berenice Michels, National Institute for Curriculum Development, the Netherlands; • Mariam Mohammad Ahmed, Evaluation Institute, Qatar; • Gabriela Noveanu, Institute for Educational Sciences, Romania; • Galina Kovaleva, Russian Academy of Education, Russian Federation;

	• Maria Pilar Jimenez Aleixandre, Universidad de Santiago de Compostela, Spain ; • Wolfgang Dietrich, National Agency for Education, Sweden ; • Gerry Wheeler, United States The TIMSS National Research Coordinators (NRCs) work with international project staff to ensure that the study is responsive to their concerns, both policy-oriented and practical, and are responsible for implementing the study in their countries.
Data set basic information	
Dates.	1995, 1999, 2003, 2007, and 2011
Countries.	In TIMSS 2011, 63 countries and 14 benchmarking participants participated:, Armenia, Australia, Austria, Bahrain, Belgium (Flemish), Botswana, Chile, Chinese Taipei, Croatia, Czech Republic, Denmark, England, Finland, Germany, Ghana, Honduras, Hong Kong SAR, Hungary, Indonesia, Iran, Israel, Italy, Japan, Jordan, Kazakhstan, Korea, Kuwait, Lebanon, Lithuania, Macedonia, Malaysia, Malta, Morocco, Netherlands, New Zealand, Northern Ireland, Norway, Oman, Palestinian National Authority, Poland, Portugal, Qatar, Romania, Russian Federation, Saudi Arabia, Serbia, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sweden, Syrian Arab Republic, Thailand, Tunisia, Turkey, Ukraine, United Arab Emirates, United States, Republic of Yemen. Benchmarking Participants: Abu Dhabi, UAE; Alabama, USA; Alberta, Canada; California, USA; Connecticut, USA; Dubai, UAE; Florida, USA; Indiana, USA; Massachusetts, USA; Minnesota, USA; North Carolina, USA; Ontario, Canada; Quebec, Canada.
Contents.	The TIMSS international assessment of student achievement is comprised of written tests in mathematics and science together with a set of questionnaires that gather information on the educational contexts in which students learn mathematics and science. It assesses student performance in the major content domains of the subjects (i.e., number, algebra, geometry, and data and chance in mathematics at the eighth grade and biology, chemistry, physics, and earth science in eighth-grade science) as well as performance in mathematics and science overall. TIMSS will also administer background questionnaires to curriculum specialists, the students in participating schools, their mathematics and science teachers, their school principals, and their parents. The questions are designed to measure key elements of the curriculum as it is intended, as it is implemented, and as it is learned.
Core data includes:	Core data includes data on mathematics and science (curricula measures). In mathematics at the fourth grade level there are three content domains in TIMSS 2011: Number, Geometric Shapes and Measurement, and Data Display. Eighth-grade math in TIMSS 2011 has four content domains: Number, Algebra, Geometry, Data and Chance. In science, at the fourth grade level, TIMSS 2011 has three content domain scales: Life Science, Physical Science, and Earth Science. In eighth grade, the content domains are Biology, Chemistry, Physics and Earth Science.
Contextual data includes:	The background questionnaires are administered to curriculum specialists, to students in participating schools, their mathematics and science teachers, their school principals, and their parents. The background questionnaires focus on elements of the curriculum to identify intentions for the curriculum, the manner in which it is implemented, and how it is learned by the students. Other data on classroom and school environment are also collected, as well as socio demographic information of the students.
Requirements of access.	Access to TIMSS data available for free at http://timss.bc.edu/
Methodology	
Sources and collection methods.	Data was collected through the administration of background questionnaires to pupils, their teachers and school principals, and assessment questionnaires to pupils.
Unit of analysis	Fourth and eighth grade students in participating countries are principal units of analysis. Schools and classroom are potentially units of analysis also. Target populations were the grades with the largest proportions of 9 year olds and 14 year olds in each participating country.
What is the sample design?	Probability proportionate to size sampling is used to select the schools. Within schools a class group is randomly sampled to take the tests (one mathematics class and one science class). To sample which students take which questions, TIMSS 2011uses a matrix-sampling approach that involves

	packaging the entire assessment pool of mathematics and science questions into a set of 14 student achievement booklets, with each student completing just one booklet. Each question, or item, appears in two booklets, providing a mechanism for linking together the student responses from the various booklets. Booklets are distributed among students in participating classrooms so that the groups of students completing each booklet are approximately equivalent in terms of student ability. Using Item-Response Theory (IRT) scaling techniques, a comprehensive picture of the achievement of the entire student population is assembled from the combined responses of individual students to the booklets they are assigned. This approach reduces to manageable proportions what otherwise would be an impossible student burden, albeit at the cost of greater complexity in booklet assembly, data collection, and data analysis.
Sample threshold	The national threshold under which replacement schools are included, or the threshold under which comparability is compromised, is 85%
Collection window?	Data collection for the 2011 survey was carried out in October-December 2010 for southern hemisphere countries, and in March-June 2011 for northern hemisphere countries.
Delivery in the field	
Training and management of field staff	The TIMSS & PIRLS International Study Center at Boston College serves as the International Study Center for IEA's studies in mathematics, science, and reading – the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). The staff at the Study Center is responsible for the design and implementation of the study. In each country, a national representative, called the National Research Coordinator (NRC), is responsible for implementing PIRLS in accordance with international procedures.
Translations and other procedures?	Translation methods are described online at http://timssandpirls.bc.edu/methods/pdf/TP_Translation_Verif.pdf
Lengths of tests	In the 2011 survey, for fourth graders, the exam consisted of two 36 minute test sections and one thirty minute questionnaire, separated by breaks; for eight graders, the exam consisted of two 45 minute test sections and one thirty minute questionnaire, separated by breaks.
Replacement sampling	
What modules are included?	To facilitate the process of creating the student achievement booklets, TIMSS groups the assessment items into a series of item blocks, with approximately 10-14 items in each block at the fourth grade level, and 12-18 items for eighth graders. As in the TIMSS 2007 assessment, the 2011 survey had a total of 28 blocks, 14 for math and 14 for science. 8 of the 14 blocks for each discipline were used for understanding trends, and the remaining 12blocks were released into the public domain for research purposes. Accordingly, the 2011 TIMSS had 16 blocks for trend items, and 12 blocks of new items. Contents of TIMSS include questionnaires assessing students' knowledge and understanding of mathematics and science through a range of questions in each subject. Two question formats are used in the TIMSS assessment – multiple-choice and constructed-response. At least half of the total number of points represented by all the questions will come from multiple-choice questions. Each multiple-choice question is worth one score point. Back ground questionnaires are administered to curriculum specialists, and to the students in participating schools, their mathematics and science teachers, and their school principals. The questions are designed to measure key elements of the curriculum as it is intended, as it is implemented, and as it is learned.
Can countries add their own items?	Yes, countries are able add questions.
Next wave/project	
Next date and availability.	2015
New content / questions.	N/A
Countries to be added.	N/A
Previous content not repeated in the next wave.	N/A
Countries not repeating the study in the next wave.	N/A
Links	

Home page for the website	http://timss.bc.edu/
To access the data online	The international database for TIMSS 2003 data is available online at http://timss.bc.edu/timss2003i/userguide.html The international database for TIMSS 2007 data is available online at: http://timss.bc.edu/TIMSS2007/index.html The international database for TIMSS 2011 data is available online at: http://timssandpirls.bc.edu/timss2011/index.html
To access the international report	Aggregate country data is available in the PIRLS 2001 report online at http://timss.bc.edu/pirls2001i/PIRLS2001_Pubs_IR.html. Aggregate country data is available in the PIRLS 2007 report online at: http://timss.bc.edu/TIMSS2007/index.html The international report for TIMSS 2011 data is available online at: http://timssandpirls.bc.edu/timss2011/index.html
How to access the technical reports	The TIMSS 2003 technical report is available online at http://timss.bc.edu/timss2003i/technicalD.html The TIMSS 2007 technical report is available online at: http://timss.bc.edu/TIMSS2007/PDF/T07_AF.pdf The TIMSS 2011 technical report is available online at: http://timssandpirls.bc.edu/timss2011/frameworks.html
To access national reports	Refer to the TIMSS Encyclopaedia for each survey year, available online at: http://timssandpirls.bc.edu/#
To access a list of research that uses the data	Research based on TIMSS data available online at http://timss.bc.edu/isc/publications.html
Contact email:	isc@bc.edu

ANNEX 3: HOUSEHOLD SURVEYS

EQLS – European Quality of Life Survey.

Origin of the study	The enlargement of the European Union since May 2004 has included 12 new Member States. This puts diversity at the forefront of the European Union – diversity in living conditions, in cultural traditions, and outlook. Nurturing this cultural diversity is at the very heart of the European ideal. But large differences in material resources and living standards, in political participation rates, in levels of trust in public institutions and in how needs and responsibilities are understood, can lead to tensions and conflict. To help foster cohesion in this larger and more diversified Union, policymakers and civil society actors need to draw on precise information about how people live and how they perceive their circumstances. Against this background, the European Foundation for the Improvement of Living and Working Conditions launched its first ever pan-European Quality of life survey (EQLS) in the summer of 2003.
Management Structure	
Statement of purpose.	To gather information for comparisons between countries as well as between demographic, social and economic groups, the EQLS report documents material conditions, employment situations, living and working conditions, family and community life, health and housing in the participating countries. EQLS looks at the views of Europe's citizens on living conditions, their subjective well-being and their assessments of the society in which they live. More in-depth analyses on specific issues raised in the survey will form a key part of the Foundation's ongoing series of reports on Quality of Life in Europe. This report provides a unique insight into Europeans' quality of life today aimed at providing a useful contribution towards shaping the policies which seek to improve living and working conditions throughout Europe.
Funded by	European Foundation for the Improvement of Living and Working Conditions; it is funded through the general budget of the European Commission and the allocation of funds to Eurofound is decided in the official budgetary process (between the Commission, European Parliament and the Council of Ministers)
Affiliations.	In 2007, the fieldwork was organised and coordinated by GfK EU3C, in cooperation with national fieldwork partners in each country.
Government partners	
Expert panels	**The EQLS questionnaire is developed by an expert panel. For example, in 2003, the interview questionnaire was de**veloped by a research consortium consisting of academics from: • Institute of Philosophy and Sociology, PAS Warsaw, Nowy Swiet 7200330, Warsaw, Poland; • Dept of Sociology, Middle East Technical University, 06531 Ankara, Turkey; • Baltic Surveys, Sermuksniu 6º, 2001 Vilnius, Lithuania; • Centre for Social and Economic Strategies, Faculty of Social Sciences, Charles University, Celetná 20, 116 36 Praha 1, Czech Republic; • FIF UK, Gondova, 281801 Bratislava, Slovakia; • Central Statistical Bureau of Latvia 1, Lacplesa Str., Riga, LV-1301, Latvia; • Casa Gizimin, Pres A, Buttigieg Street, Zejtun ZTN 03, Malta; • Hungarian Central Statistical Office, Angol u. 77, Budapest, Hungary; • Economics Research Centre, University of Cyprus, P.O.Box 20537, 1678 Nicosia, Cyprus; • Institute of Economics, Bulgarian Academy of Sciences, 3 Aksakov Str., 1040 Sofia, Bulgaria; • Saar Poll Ltd, Veetorni 4, Tallinn 10119, Estonia; • Faculty of Social Sciences, University of Ljubljana, Kardeljeva ploščad 5, P.O. Box 2547, 1001 Ljubljana, Slovenia; • Universitatea Bucuresti, Institutul de Cercetare a Calitatii Vietii, Calea 13 Septemrie, No. 13, Bucuresti, Romania.
Data set basic information.	
Dates.	2003,2007 and 2011
Countries.	EQSL 2003, 28 Countries: Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France,

	Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Turkey and United Kingdom. EQSL 2007, 31 Countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Turkey and United Kingdom. EQSL 2011, 34 countries: EU 27 and Croatia, Iceland, FYR Macedonia, Montenegro, Kosovo, Serbia and Turkey
Contents.	The EQLS represents an ambitious attempt to explore quality of life in a wide range of countries. EQLS therefore examines a range of issues, such as employment, income, education, housing, family, health, work-life balance, life satisfaction, perceived quality of society, attitudes towards migrants, public services, and community participation. The 2011 questionnaire includes 183 question items, conducted face to face.
Core data includes:	The core focus of the survey is on employment, economic resources, family and households, community life and social participation, health and health care, knowledge, education and training.
Contextual data includes:	Socio-demographic information such as: household size, employment status, level of education and marital status
Requirements of access.	Available through the Data Archive. Registration is required and standard conditions of use apply (including providing information about usage). For more details see http://www.eurofound.europa.eu/surveys/availability/index.htm
Methodology	
Sources and collection methods.	The EQLS uses face-to-face interviews (in people's homes) for data collection. Computer-assisted personal interviews were used in 20 countries, paper and pencil interviews were used in 7 countries in 2011 wave.
Unit of analysis.	For EQLS 2011; Adults (aged 18 years and over) resident in and speak the language of one of the 27 European Union countries, plus Croatia, Iceland, FYR Macedonia, Montenegro, Kosovo, Serbia and Turkey
What is the sample design	For EQLS 2003 a multi-stage stratified random sample was used. For larger countries around 1000 face-to-face interviews took place, in smaller countries (including Cyprus, Estonia, Luxembourg, Malta and Slovenia) around 600 interviews were conducted. In all totalling 26,000 interviews. Sampling points were drawn after stratification by region and degree of urbanisation. The random route method was used in the majority of countries (see links below for further details). The specifications for this method were that one address had to be drawn at random in each of the selected sampling points. That starting address formed the first of a cluster of a maximum of 20 addresses. The remainder of the cluster was selected by standard random route procedure from the initial address. Weighting variables were produced according to age, gender and region within country. In addition, several weighting variables were also produced that adapt the sample size of each country to the proportion of their population within several recognised EU country groupings, e.g. EU15, EU25, and AC10. For EQLS 2007 the same sampling methodology and sample size was used, with the exception of Germany and Turkey, where 2000 interviews are completed in each country. For EQLS 2011, a target number of 1000 interviews was set for most of the countries. The seven biggest member countries had bigger sample sizes bigger than those in the previous wave of the EQLS
Sample threshold	
Collection window	For EQLS 2011 interviews were carried out face to face in people's homes, during September 2011 and December 2011. Computer-assisted personal interviews were used in 20 countries, paper and pencil interviews were used in 7 countries in the 2011 wave.
Delivery in the field	
Training and management of the field staff	The fieldwork was carried out by TNS-Opinion in 2007 and by GfK EU3C in 2011. A network of national contractors, who were provided additional training specific to the survey, carried out the data collection in each country.
Translation and other procedures	Altogether, the questionnaire was translated into 25 languages. For languages used in more than one country, adapted versions of the questionnaire were produced to account for country-specific use of a common language.
Replacement sampling	Sampling is done without replacements
What modules are included?	No modules. One standard questionnaire focussing on employment, economic resources, family and households, community life and social participation, health and health care, knowledge, education and training.

Can countries add their own items?	No
Next wave/project.	
Next date and availability.	2011 is the latest survey; the Foundation published first results in November 2011. Results from candidate of pre-accession countries will be made available in 2013,
New content / questions.	The content remains the mainly same and the questionnaire repeated many of the questions used in 2003 and 2007 waves but introduced some questions. Out of 183 questions in the 2011 questionnaire, 30% are trend questions available in all three waves, 50% are available in 2007 and 2011 and 50% are new items.
Countries to be added.	Croatia, Macedonia and Norway were added to 2007 survey, Iceland, Kosovo, Montenegro and Serbia were added in the 2011 survey.
Previous content not repeated in the next wave.	Recommendations to improve the methodology are reflected in the 2011 survey. Eurofound reviewed the questionnaire in the context of policy debates and increasing expertise in quality of life research and monitoring of subjective well-being.
Links	
Home page for the website	http://www.eurofound.europa.eu/areas/qualityoflife/eqls/index.htm
To access the data online	The 2011 data is available at http://www.eurofound.europa.eu/surveys/eqls/2011/index.htm
To access the international report	The EQSL 2003 international report is available online at http://www.eurofound.europa.eu/pubdocs/2004/105/en/1/ef04105en.pdf The EQSL 2007 international report is available online at: http://www.eurofound.europa.eu/pubdocs/2009/02/en/2/EF0902EN.pdf The EQSL 2007 international report is available online at: http://www.eurofound.europa.eu/publications/htmlfiles/ef1264.htm
How to access the technical reports	Research based on EQSL 2003 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/2003/eqlsfindings.htm Research based on EQSL 2003 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/eqls2007/results.htm
To access national reports	Research based on EQSL 2003 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/2003/eqlsfindings.htm Research based on EQSL 2007 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/eqls2007/results.htm
To access a list of research that uses the data	Research based on EQSL 2003 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/2003/eqlsfindings.htm Research based on EQSL 2007 data available online at http://www.eurofound.europa.eu/areas/qualityoflife/eqls/eqls2007/results.htm
Contact email:	postmaster@eurofound.europa.eu

ESS – European Social Survey.

Origin of the study	The ESS was initiated and seed-funded by the European Science Foundation, the body representing most of Europe's main national academic funding agencies. Prompting their decision was their realization that most cross-national attitude surveys in Europe at that time were not of sufficient rigour to draw on as reliable sources of knowledge about value change in Europe. An Expert Group was created to look into the possibility of starting a new time series on value change that would be an example of methodological and substantive rigour. The Expert Group report was enthusiastic, with the result that two new Committees were then appointed by the European Science Foundation to pursue the matter – a Steering Committee and a smaller Methodology Committee which was charged with crafting the structure and design of the proposed new time series. The two committees produced a joint 'Blueprint', which was subsequently used to form a more detailed application by six of the seven institutions that still form the ESS Central Coordinating team. The application was successful. However, one final hurdle still needed to be cleared. The EC contribution covered only the project's central design and coordination. The bulk of the funding – for fieldwork and coordination at a national level - was intended to come from national academic research councils in each participating country. As it turned out, 22 countries in Round 1, 26 in Round 2, and 25 in Round 3 agreed to produce the necessary funding, together with a commitment to work within a rigorous centrally-determined specification, and the ESS came to being in 2001. The first survey was fielded in 2002/03, the second in 2004/05, the third in 2006/07, the fourth in 2008/2009, and the fifth in 2010/2011.
Management Structure	
Statement of purpose	To gather data about changing values, attitudes, attributes and behaviour patterns within European polities. Academically driven, but designed to feed into key European policy debates, the ESS is designed to measure and explain how people's social values, cultural norms and behaviour patterns are distributed, the way in which they differ within and between nations, and the direction and speed at which they are changing.
Funded by	The project is funded jointly by the European Commission, the European Science Foundation, and academic funding bodies in each participating country.
Affiliations	
Co-ordinators	**Central Coordination Team** The following people in each of the following institutions are responsible for the design and coordination of the European Social Survey. The lead institution is the Centre for Comparative Social Surveys, City University, London, UK
Government partners	European Commission
Expert panels	**Scientific Advisory Board** The multi-national Scientific Advisory Board is chaired by Professor Max Kaase, It is comprised of one representative selected by each national funding agency, a network of National Coordinators representing each country, a small expert multii-nation Methods Group, a Funders' Forum representing all countries, and representatives from the European Commission and the European Science Foundation, respectively. **Methods Group** This small, multi-national methods group advises the CCT on technical and methodological aspects of the survey. Its membership is as follows: • Denise Lievesley, Chair • Norman Bradburn, US National Science Foundation • Paolo Garonna, UNECE • Lars Lyberg, Statistics Sweden • Vasja Vehovar, University of Ljubljana • Peter Lynn, University of Essex **The sampling panel** This team of specialists advises on specific issues regarding sampling. They are responsible for advising National Coordinators on sample selection and will finally 'sign off' all national sample designs. The 2010 Sampling Panel was chaired by Matthias Ganninger, GESIS. **Translation taskforce** A team of translation specialists is responsible for guiding the translation process. They have provided guidance notes for and advice to National Co-ordinators, to enable them to work effectively with those who perform the

	translation from the source questionnaire in each country. **Questionnaire Design Team on Trust in Police and Courts** This questionnaire design team was chaired by Jonathon Jackson, London School of Economics and Political Science, UK. **Questionnaire Design Team on Work, Family, and Wellbeing** This questionnaire was designed by Duncan Gallie, University of Oxford, UK.
Data set basic information.	
Dates.	2002/2003, 2004/2005, and 2006/2007, 2008/2009, and 2010/2011.
Countries.	Round 1: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, Switzerland, UK Round 2: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, UK, Ukraine Round 3: Austria, Belgium, Bulgaria, Cyprus, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Latvia, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, UK, Ukraine Round 4: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, UK, Ukraine Round 5: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, UK, Ukraine
Contents.	The ESS questionnaire includes two main sections: a 'core' module and two or more 'rotating' modules, repeated at intervals (see below). The fourth core module carries what it calls the “most comprehensive set of socio-structural (background) variables of any cross-national survey.” In addition, a supplementary questionnaire is given to respondents at the end of the main interview. The first part of this questionnaire is a human values scale (part of the core), while the second is devoted to measures to help evaluate the reliability and validity of items in the main questionnaire.
Core data includes:	The core module aims to monitor change and continuity in a wide range of social variables, including: media use, social and public trust; political interest and participation; socio-political orientations, governance and efficacy; moral, political and social values; social exclusion, national, ethnic and religious allegiances; well-being, health and security; demographics and socio-economics.
Contextual data includes:	The socio-demographic profile is comprised of a series of questions on household composition; sex; age; marital status; type of education and occupation of respondent, partner, and parents; union membership; and income.
Requirements of access.	Access to ESS data available online: http://www.europeansocialsurvey.org/ess.nsd.uib.no/ess
Methodology	
Sources and collection methods.	Data collection takes place every two years, by means of face to face interviews of around an hour in duration, followed by a short supplement.
Unit of analysis.	People 15 years and older, with no upper age limit, who are resident in the country, regardless of nationality, citizenship or legal status.
What is the sample design?	The sample is selected using random (probability) samples, with comparable estimates based on full coverage of eligible residential populations fifteen years of age and older.
Sample threshold	Information on sampling can be found at http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=80&Itemid=653
Collection window	Between September and December.
Delivery in the field	
Training and management of	The Central Coordination Team (CCT) is responsible for the design and coordination of the European Social Survey

the staff	A structure has been adopted that combines central co-ordination with participation from all responding countries, and independent advice from teams of independent academics (see *Expert panels* above).
Translation and other procedures	Translation is completed under the supervision of the CCT approximately 18 months before the fieldwork is completed.
Length of test	About an hour, plus supplement.
Replacement sampling	Information on sampling can be found at http://www.europeansocialsurvey.org/index.php?option=com_content&view=article&id=80&Itemid=653
What modules are included?	The questionnaire includes two main sections, each consisting of approximately 120 items; a 'core' module which will remain relatively constant from round to round, plus two or more 'rotating' modules, repeated at intervals. In 2005/06 the two rotating modules were: Personal & Social Well-being: Creating indicators for a flourishing Europe, and The Timing of Life: The organisation of the life course in Europe. Each contained 50 additional items.
Can countries add their own items?	No.
Next wave/project.	
Next date and availability.	The sixth round is underway.
New content / questions.	
Countries to be added.	To be confirmed.
Previous content not repeated in the next wave.	To be confirmed.
Countries not repeating the study in the next wave.	To be confirmed.
Links	
Home page for the website	http://www.europeansocialsurvey.org
To access the data online	Via homepage.
To access the international report	Via homepage.
How to access the technical reports	Via homepage.
To access national reports	
To access a list of research that uses the data	Research based on ESS data available online at http://www.europeansocialsurvey.org/index.php?option=com_content&task=view&id=110&Itemid=148.
Contact email:	Data queries can be addressed to essdatasupport@nsd.uib.no. **Error! Hyperlink reference not valid.**

EU SILC – Statistics on Income and Living Conditions.

Origin of the study	The European project on Statistics on Income and Living Conditions continues the work of the pioneering European Community Household Panel survey (ECHP). THE ECHP ran from 1994to 2001. Since that time, the European political scene has changed. First, there was the introduction of an open method of coordination in the fields of social inclusion and pension reform. Second, the number of EU member states increased. And finally, the United Nations called for new recommendations on the collection of household income statistics In recognition of these changes, the ECHP was progressively replaced with data collection under the EU-SILC (Statistics on Income and Living Conditions). Seven countries launched a preliminary version of EU-SILC in 2003. The project was formally launched in 2004, and in 2005 the statistical information was gathered for the first time.
Management Structure	
Statement of purpose.	EU-SILC is the EU reference source for comparative statistics on income distribution, living conditions and social exclusion at European level. The purpose of SILC is to allow the Member States and the European Commission to monitor national and EU progress towards key EU objectives in the area of social inclusion and of social protection, and to support mutual learning and identification of good (and bad) practices in terms of policies and institutional processes. This represents a major step forward in the development of EU cooperation in social policy, and has the potential to transform the ways in which Member States develop their national (and sub-national) policies to tackle poverty and social exclusion.
Funded by	European Commission
Affiliations.	
Co-ordinators	EUROSTAT and the European Commission.
Government partners?	EUROSTAT harmonizes and distributes data collected by the national statistics institutes of member countries.
Expert panels	The development of indicators, under the responsibility of the Social Protection Committee (SPC) Indicators Sub-Group since February 2001, is a dynamic process. The work of the national delegations of experts, who make up the Group, and the secretariat provided by the European Commission Directorate-General (in close cooperation with EUROSTAT), has allowed the set of indicators (and breakdowns of these) to be considerably enriched.
Data set basic information.	
Dates.	2004, 2005, 2006, 2007, 2008, 2009, 2010, and onwards (yearly).
Countries.	For SILC 2005: EU member countries (minus Romania and Bulgaria, which became member countries in 2007), plus Norway and Iceland. For SILC 2006: EU member countries, Norway and Iceland For SILC 2007 onward: EU member countries, Norway, Iceland, Bulgaria, Romania, Turkey, and Switzerland. Ongoing updates on participant countries can be found at http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc
Contents.	EU-SILC is a multi-dimensional instrument focused on income but covering at the same time housing, labour, health, demography, education so as to allow studying the multidimensional approach of social exclusion. It is composed of primary (annual) and secondary (module) target variables. Given the principle of flexibility of the implementation of the SILC project at national level, the corresponding sequence of questions needed to construct one target variable may vary from one country to another. The primary target variables are either household or individual (for persons aged 16 and more) information and are regrouped into domains: • At household level, five domains are covered ((1) basic/core data, (2) income, (3) housing, (4) social exclusion, (5) labour information. • The personal level is regrouped into six domains ((1) basic/demographic data, (2) income, (3) education, (5) labour information and (6) health). For countries using the integrated design, all variables will be in both cross-sectional and longitudinal components. The secondary target variables are introduced every four years or less frequently. One module per year is included in the cross-sectional component. The first EU-SILC modules are relating to:

	• 2005: Inter-generational transmission of poverty; • 2006: Social participation • 2007: Housing conditions • 2008: Over-indebtedness/Financial exclusion • 2009: Deprivation • 2010: Intra-household sharing of resources
Core data includes:	Data on income, poverty, social exclusion, and living conditions.
Contextual data includes:	Socio-demographic contextual data is an integrated part of the core data/questionnaire.
Requirements of access.	The EU-SILC data are cleaned and imputed, and then de-identified individual records are transmitted to EUROSTAT. Information on how to access EU-SILC data is available at http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/documents/EN-EU-SILC-MICRODATA.pdf EU-SILC frameworks provide for the release of anonymised micro data to researchers (encrypted CD-ROM with documentation).
Methodology	
Sources and collection methods.	EU-SILC framework fosters the use of existing sources and/or administrative data. However, in practice not all EU-SILC variables can be obtained from register and administrative data. So, two groups of countries can be identified based on where data is sourced for EU-SILC. The first group, the so called register countries (DK, FI, IS, NL, NO, SE, SI), are countries where most income components and some demographic information are obtained through administrative registers. Other personal variables are obtained through interview. In the remaining countries, except Ireland, the full information is obtained through survey among household and interview with household members. In Ireland, upon the explicit agreement of the household collected, the information is obtained from administrative information. The specific mode of collecting information also varies from country to country. PAPI (Performance Application Programming Interface) is still the main collection mode. CAPI (Computer Assisted Interviewing) is implemented in some countries. Germany is the only country where questionnaires were sent by post. The editing facility provided by computer assisted mode is implemented in about half of the countries has a positive impact on the quality of the micro-data collected and reduces costs of data collection and edition.
Unit of analysis.	In EU-SILC two types of collection of household and individual variables have been allowed. In most countries (the non-register countries), all members aged 16 or higher of selected households have been asked to fill in a personal questionnaire. In the register countries (DK, FI, IS, NL, NO, SE, SI), only a selected household respondent receives a personal questionnaire and household and income variables are collected either through register or through the selected respondent.
What is the sample design?	According to the Commission Regulation on sampling and tracing rules, for all components of EU-SILC (whether survey or register based), the cross-sectional and longitudinal (initial sample) data are to be based on a nationally representative probability sample of the population residing in private households within the country, irrespective of language or nationality. Regulation N° 1177/2003 defines the minimum effective sample sizes to be achieved. As described on the EU-SILC website, The minimum size of the sample of the overall population which is surveyed every year is: • Cross-sectional data operation: about 130,000 households and 270,000 persons aged 16 and more are interviewed in the European Union countries. • Longitudinal data operation: about 100,000 households and 200,000 persons aged 16 and more are interviewed in the European Union countries.
Sample threshold	Sample sizes are defined at http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/documents/tab/Tab/EU-SILC%20sample%20size.pdf
Collection window	National surveys also differ in the period during which the fieldwork is carried out. Regulation recommends that the one shot survey fieldwork extends over less than 4 consecutive months and the lag between income reference period and fieldwork is limited to 8 months. When continuous surveys are used, the sample allocation over time is controlled and weighting adapted to produce unbiased estimates of the annual average.
Delivery in the field	
Training and management of the field staff	Varies by country.

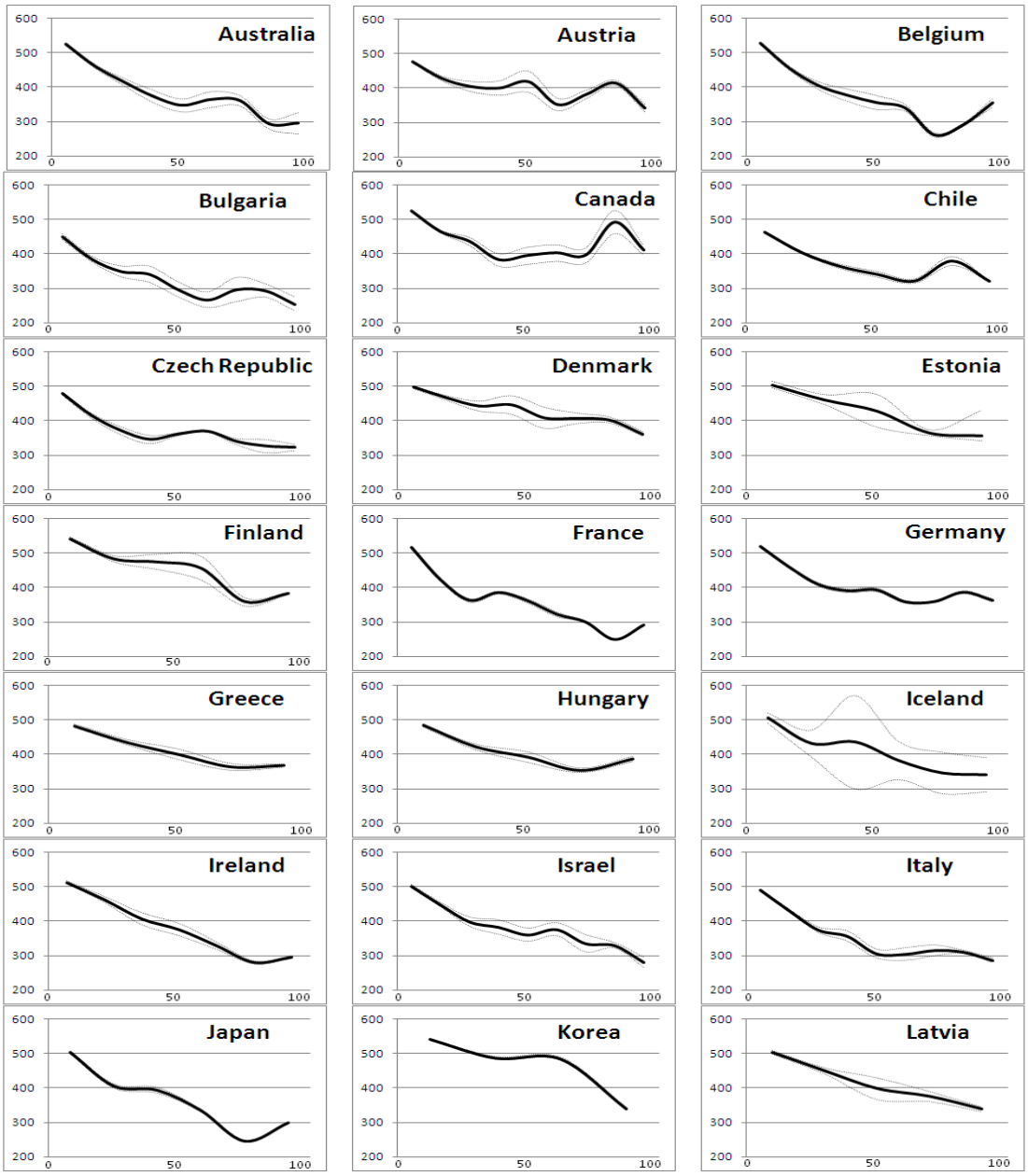
Translation and other procedures	Varies cy country.
Length of tests	Varies by country.
Replacement sampling	Additional households (and individuals) for a given cross sectional wave are sampled without replacement.
What modules are included?	SILC collects information on income distribution, living conditions and social exclusion by means of two sets of indictors: overarching indicators and social inclusion. However to anchor EU-SILC in the National Statistical System, survey design is flexible. In this way, the cross-sectional and longitudinal data can come from separate sources, i.e. the longitudinal dataset does not need to be "linkable" with the cross-sectional dataset at the micro-level. Of course, such linkage was not precluded, and is actually frequently met because the two types of data come from the same source. Depending on the country, micro-data could come from: • Two or more national sources (surveys and/or registers). • One or more existing national sources combined or not with a new survey. • A new harmonized survey to meet all EU-SILC requirements. An integrated design ('the rotational design') for those countries that launched a new survey was proposed by EUROSTAT and implemented by countries. Rotational design refers to the sample selection based on a number of sub-samples or replications, each of them similar in size and design and representative of the whole population. From one year to the next, some replications are retained, while others are dropped and replaced by new replications.
Can countries add their own items?	Yes. In most countries, EU-SILC variables are embedded in a larger national survey.
Next wave/project.	
Next date and availability.	Results are published annually. Next release date TBA.
New content / questions.	Varies by country.
Countries to be added.	New member states are added if accession completed before start of survey
Previous content not repeated in the next wave.	None
Countries not repeating the study in the next wave.	None.
Previous waves/projects.	Over an eight year period (from 1994 to 2001), the ECHP (European Community Household Panel), ran in 14 of the then 15 Member States (with the exception of Sweden), and served as the source for many of the commonly agreed social inclusion indicators for this period. The role of the ECHP has therefore been crucial for the first two rounds of EU-15 National Action Plans on inclusion (2001 and 2003). In parallel, international recommendations on income (Expert Group on Household Income Statistics (The Canberra Group), 2001) were developed, and collection of gross income at component level (and not net income as implemented in the ECHP) appeared to be preferable for income distribution analysis. In order to solve the ECHP technical problems, to conform to the internationally agreed definition of income, and to extend data collection to the enlarged EU (and beyond), the decision was taken to stop the ECHP and launch EU-SILC.
Links	
Home page for the website	http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc
To access the data online	Due to the personal information contained in EU-SILC microdata, direct access to anonymised data is only provided by way of research contracts to universities, research institutes, national statistical institutes, central banks, and the European Central Bank. Further information on accessing EU-SILC data can be found at http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/documents/EN-EU-SILC-MICRODATA.pdf If an organization decides to order the data set, an official access request must be made.
To access the international report	EU-SILC publications are available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc/publications

How to access the technical reports	See homepage.
To access national reports	
To access a list of research that uses the data	Research based on SILC data is available online at http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1913,47567825,1913_58814988&_dad=portal&_schema=PORTAL#C
Contact email:	estat-microdata-access@ec.europa.eu

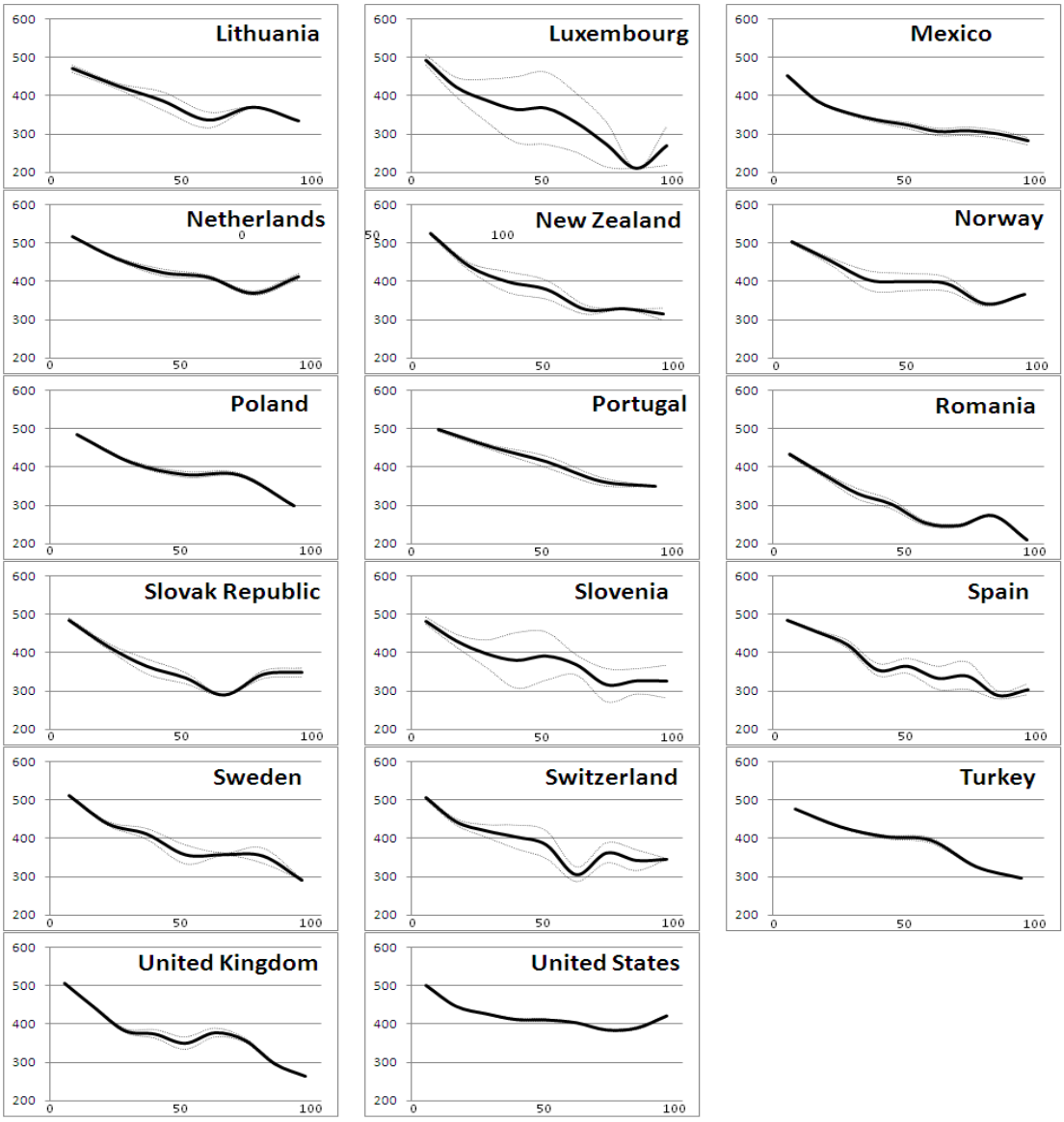
ANNEX 4. TABLES AND CHARTS FOR PISA EVALUATION

A4.1 Country level associations between missing values count and reading literacy achievement, 2009

Trends move along the standardised x-axes from left to right in order of low or no missing values to high number of missing values (x-axis is the proportion of missing responses, capped at 100 unique values).



A4.1 Country level associations between missing values count and reading literacy achievement, 2009 (cont.)



Source: OECD calculations of OECD PISA data 2009. Solid lines represent mean reading literacy scores, dotted lines are plus and minus one standard error of the mean.

A4.2 Logistic regression beta coefficients for membership of missing group for educational deprivation 2009

	Unweighted data			Weighted data		
	Sex	Migrant status	Parent edu.	Sex	Migrant status	Parent edu.
Australia	0.001	0.013*	-0.001	-0.001	0.006***	0.002***
Austria	-0.005	0.011	0.003	-0.003**	0.001	0.001*
Belgium	0.006	-0.002	0.007***	0.008***	-0.006***	0.006***
Bulgaria	-0.008	0.05***	0.02***	-0.007***	0.043***	0.026***
Canada	-0.002	0.009***	0.004**	-0.002***	0.004***	0.005***
Chile	0.003	-0.023*	0.001	0.002**	-0.022***	0.000
Czech Republic	-0.009***	0.05***	-0.001	-0.008***	0.042***	0.000
Denmark	-0.003	0.003	0.009***	-0.004***	0.001	0.006***
Estonia	0.007**	0.005	0.022***	0.01***	0.002	0.018***
Finland	0.000	-0.001	-0.006*	0.002**	-0.006**	-0.005***
France	-0.002	0.004	0.004**	-0.003***	0.007***	0.004***
Germany	-0.009*	0.048***	0.001	-0.009***	0.043***	0.002***
Greece	0.011**	0.017*	-0.003	0.008***	0.004*	0.006***
Hungary	-0.004	-0.005	-0.003	-0.006***	-0.009***	-0.002***
Iceland	-0.003	0.004	-0.015***	-0.003	0.002	-0.016***
Ireland	0.004	0.028	0.006	0.005**	0.022***	0.009***
Israel	0.003	0.051***	0.007***	0.002*	0.046***	0.008***
Italy	0.003	0.006	-0.001	0.002***	0.00	-0.002***
Japan	-0.002	0.003	0.009*	-0.002***	0.001	0.009***
Korea	0.01***	-0.004	-0.001	0.011***	-0.009***	-0.004***
Latvia	-0.006	0.002	0.001	-0.009***	0.000	0.006***
Lithuania	-0.006	0.041***	0.012***	-0.005***	0.037***	0.013***
Luxembourg	0.008*	0.049***	0.01***	0.008*	0.046***	0.01***
Mexico	0.001	0.023***	-0.003***	-0.002***	0.026***	-0.006***
Netherlands	0.001	0.036***	0.001	0.003***	0.042***	0.001
New Zealand	0.004	0.013	0.003*	0.004***	0.009**	0.003***
Norway	0.000	0.033**	0.000	-0.002	0.028***	0.000
Poland	-0.005	-0.007	-0.012***	-0.004***	-0.007***	-0.012***
Portugal	-0.005*	-0.004	-0.007***	-0.004***	-0.004**	-0.008***
Romania	0.02***	0.056***	0.007	0.017***	0.051***	0.008***
Slovak Republic	0.011**	0.033*	0.011**	0.01***	0.024***	0.009***
Slovenia	0.000	0.028***	-0.001	-0.001	0.014***	-0.001
Spain	0.000	0.006	-0.001	-0.001**	-0.003***	-0.003***
Sweden	-0.007	0.003	0.014***	-0.007***	0.002	0.011***
Switzerland	0.001	0.018***	0.004***	-0.001	0.017***	0.003***
Turkey	0.006	-0.029	0.005**	0.003***	-0.029***	0.006***
United Kingdom	0.000	0.011	0.001	0.001	0.002*	-0.002***
United States	-0.02***	0.01	0.011***	-0.019***	0.008***	0.01***

Source: OECD calculations of OECD PISA data 2009.

Note: * p<0.05, **<p<0.01, ***p<0.001.

A4.3 Logistic regression beta coefficients for membership of missing group for employment aspirations, 2006

	Unweighted data			Weighted data		
	Sex	Migrant status	Parent edu.	Sex	Migrant status	Parent edu.
Australia	-0.343***	0.176**	0.084***	-0.339***	0.124***	0.084***
Austria	-0.222***	-0.02	0.081**	-0.23***	0.001	0.104***
Belgium	-0.017	-0.319***	0.061***	-0.014	-0.162***	0.069***
Bulgaria	-0.146**	0.208	0.018	-0.163***	0.117	0.016
Canada	-0.371***	0.011	0.092***	-0.411***	-0.072***	0.077***
Chile	-0.196***	-0.025	-0.068	-0.249***	-0.024	-0.038***
Czech Republic	-0.328***	-0.098	0.022	-0.357***	-0.021	0.041***
Denmark	0.171**	0.137	0.052*	0.171***	0.101**	0.05***
Estonia	-0.185***	-0.404*	-0.121	-0.139***	-0.44***	-0.158***
Finland	-0.589***	-0.629***	0.068*	-0.583***	-0.625***	0.067***
France	-0.455***	-0.431***	0.08***	-0.457***	-0.457***	0.078***
Germany	-0.217***	-0.313***	0.086***	-0.231***	-0.313***	0.087***
Greece	-0.732***	0.007	-0.196***	-0.713***	0.043*	-0.147***
Hungary	-0.236***	0.348	0.031	-0.259***	0.44***	0.042***
Iceland	-0.45***	-0.093	0.024	-0.459***	-0.116	0.017
Ireland	-0.292***	-0.204*	0.075*	-0.297***	-0.225***	0.072***
Israel	-0.269***	0.141	0.129***	-0.235***	0.18***	0.127***
Italy	-0.036	-0.041	0.117***	-0.048***	-0.034*	0.092***
Japan	-0.092	-0.23	0.137***	-0.123***	-0.278***	0.133***
Korea	-0.334***	18.459	0.154**	-0.334***	18.441	0.167***
Latvia	-0.573***	0.43	0.055	-0.625***	0.637***	0.084***
Lithuania	-0.223***	-0.1	0.073*	-0.215***	-0.157*	0.062***
Luxembourg	-0.406***	-0.267***	0.09***	-0.403***	-0.273***	0.091***
Mexico	-0.407***	-0.455***	-0.224***	-0.473***	-0.687***	-0.172***
Netherlands	-0.194**	-0.143	0.085***	-0.23***	-0.066**	0.078***
New Zealand	-0.473***	0.204**	0.084***	-0.474***	0.226***	0.081***
Norway	-0.196***	-0.182	0.059**	-0.189***	-0.184***	0.057***
Poland	-0.095	-0.218	0.065**	-0.141***	-0.586***	0.069***
Portugal	-0.205***	-0.252**	0.077**	-0.281***	-0.342***	0.076***
Romania	-0.148	19.085	0.003	-0.146***	19.016	0.109***
Slovak Republic	-0.278***	0.317	-0.186**	-0.267***	0.526***	-0.151***
Slovenia	-0.333***	-0.217	0.108***	-0.31***	-0.082	0.105***
Spain	-0.424***	-0.43***	0.063***	-0.523***	-0.484***	0.041***
Sweden	-0.184**	-0.221	0.065***	-0.243***	-0.235***	0.043***
Switzerland	-0.022	-0.222***	0.095***	0.021	-0.154***	0.091***
Turkey	-0.849***	0.417	-0.242***	-0.78***	0.298***	-0.268***
United Kingdom	-0.242***	-0.175	0.064***	-0.271***	-0.164***	0.066***
United States	-0.458***	-0.435***	0.089**	-0.448***	-0.48***	0.123***

Source: OECD calculations of OECD PISA data 2009.

Note: * p<0.05, **<p<0.01, ***p<0.001.

A4.4 Logistic regression beta coefficients for membership of missing group for mother's employment, 2003

	Unweighted data			Weighted data		
	Sex	Migrant status	Parent edu.	Sex	Migrant status	Parent edu.
Australia	0.483***	0.226**	0.219***	0.619***	0.314***	0.187***
Austria	0.508***	0.569***	0.14***	0.54***	0.572***	0.135***
Belgium	0.261***	0.791***	0.16***	0.182***	0.821***	0.173***
Canada	0.395***	0.511***	0.217***	0.381***	0.551***	0.247***
Czech Republic	0.572***	0.331	0.184***	0.589***	0.322***	0.202***
Denmark	0.479**	1.091***	0.193***	0.347***	1.018***	0.192***
Finland	0.492***	0.965***	0.258***	0.428***	1.242***	0.262***
France	0.073	0.822***	0.052*	0.093***	0.828***	0.05***
Germany	0.288***	0.581***	0.131***	0.297***	0.558***	0.143***
Greece	0.206**	0.208	-0.226**	0.119***	0.205***	-0.232***
Hungary	0.095	-0.174	0.215***	0.059**	0.015	0.215***
Iceland	0.678***	0.202	0.162**	0.67***	0.221	0.159***
Ireland	0.231	0.251	0.235***	0.153***	0.204**	0.226***
Italy	0.405***	0.814***	0.169***	0.457***	0.424***	0.08***
Japan	0.533***	0.83*	-0.497***	0.536***	0.808***	-0.488***
Korea	0.052	-0.131	0.144***	0.069***	-0.106	0.135***
Luxembourg	0.557***	-0.012	0.127***	0.556***	-0.012	0.128***
Mexico	-0.026	-0.191	-0.184***	0.102***	-0.457***	-0.231***
Netherlands	0.129	0.831***	0.066***	0.062***	0.724***	0.052***
New Zealand	-0.005	0.981***	0.041***	-0.029	0.984***	0.053***
Norway	0.562***	0.988***	0.162***	0.576***	0.98***	0.152***
Poland	0.368***	-18.038	-0.763***	0.36***	-18.058	-0.756***
Portugal	0.229	0.468**	0.08*	0.253***	0.479***	0.078***
Slovakia	0.276***	0.434	-0.035	0.238***	0.414***	-0.027
Spain	0.385***	0.345	0.177***	0.525***	0.395***	0.199***
Sweden	0.484***	1.088***	0.176***	0.503***	1.049***	0.182***
Switzerland	0.34***	0.789***	0.188***	0.495***	0.688***	0.186***
Turkey	0.169**	-0.798	-0.211***	-0.055***	-0.637***	-0.212***
United Kingdom	0.364***	0.153	0.204***	0.277***	0.155***	0.197***
United States	0.191**	0.642***	0.087***	0.174***	0.725***	0.103***

Source: OECD calculations of OECD PISA data 2009.

Note: * p<0.05, **<p<0.01, ***p<0.001.

A4.5 Logistic regression beta coefficients for membership of missing group for parental interaction, 2000

	Unweighted data			Weighted data		
	Sex	Migrant status	Parent edu.	Sex	Migrant status	Parent edu.
Australia	-0.406***	1.223***	-0.507***	-0.654	0.968	-0.184
Austria	-0.971***	-0.748***	-0.757***	-0.937*	-0.263	-1.395**
Belgium	-0.385***	-0.365**	0.232	-0.323	-0.827	0.426
Canada	-0.602***	0.962***	0.771***	-0.483***	0.546	0.46
Chile	-0.972***	-0.43***	-0.234***	-0.903***	-0.387	-0.383
Czech Republic	-0.425***	-1.649***	-2.455***	-0.484	-1.491	-2.84**
Denmark	-0.512***	-0.363**	-0.36	-0.336	-0.419	-0.467
Finland	-0.529**	15.189	-1.134***	-0.477	15.178	-1.194
France	-0.551***	15.884	16.187	-0.695	15.94	16.237
Germany	-0.533***	-1.895***	0.695***	-0.076	-0.826	-0.178
Greece	-0.199***	-0.44***	-0.002	-0.066	-0.312	-0.003
Hungary	-0.98***	-2.743***	14.551	-1.256	-3.262***	14.833
Iceland	-0.064	16.412	16.407	0.019	16.44	16.44
Ireland	0.196	16.008	-0.942***	0.309	15.907	-0.698
Israel	0.899***	2.35***	-0.596***	0.276	0.823	0.03
Italy	-1.345***	-0.877***	0.401***	-0.974**	-1.074	-0.07
Luxembourg	-0.196	-1.31***	-0.96**	-0.079	-1.496***	-0.745
Mexico	0.011	-0.373***	-1.433***	-0.281	-0.47	-1.124**
Netherlands	-0.821***	-1.394***	-0.293***	-1.201	-1.006	-0.703
New Zealand	-0.1	16.776	-1.528***	-0.201	16.785	-1.508*
Norway	-0.652***	-0.61**	0.365	-0.59	-0.263	-0.485
Poland	-0.312***	-1.923***	17.637	-0.171	-1.452	17.615
Portugal	-0.628***	16.817	-0.137	-0.548	16.855	-0.195
Spain	-0.807***	16.599	-1.044***	-0.468	16.45	-1.246***
Sweden	0.768***	-0.506**	-1.055***	0.736**	-0.415	-1.324
Switzerland	-0.156	-0.916***	1.832***	-0.168	-0.751*	0.494
United Kingdom	0.674***	-0.437***	-1.125***	0.052	-0.454	-0.549
United States	-0.06***	-0.128***	1.012***	0.188	-0.818	0.042

Source: OECD calculations of OECD PISA data 2009.

Note: * p<0.05, **<p<0.01, ***p<0.001.

ANNEX 5: QUESTIONNAIRE ITEMS BY DIMENSION

Socio-demographics

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Age	EQLS	2007	Household	Age of each household[1]	Scale:
	ESS	2008	15+	Age - grouped	Ordinal scale:
	ESS	2008	Household	Age of each household[1]	Scale:
	EU-SILC	2008	Household	Age[1]	
	PIRLS	2001	9 to 10	When were you born?	Tick box: Month; Year.
	PISA	2000, 2009	15	On what date were you born?	Date:
	TIMSS	2003	10	When were you born?	Month and year
Family form and household size	EQLS	2007	15+	Marital status?	Tick box: Married or living with partner; Separated or divorced and not living with partner; Widowed and not living with partner; Never married and not living partner; (Don't know/no answer)
	ESS	2008	15+	Current legal marital status?	Tick box: Married ; separated ; divorced ; widowed ; never married
	ESS	2008	15+	Have you ever lived with a partner without being married to them?	Yes/No:
	ESS	2008	15+	If respondent single then? Divorced / never married?	Yes/No:
	ESS	2008	15+	Is respondent married / in a civil partnership / living with partner?	Yes/No:
	ESS	2008	Household	Relationship to each household[1]	Tick box: spouse/partner; son/daughter; parent/step-parent/parent-in-law; daughter/son-in-law; grandchild; brother/sister; other relative; other non-relative
	CIVED	1999	14	Altogether, how many people live in your home?	Scale:
	EQLS	2007	15+	How many children in the house?	Scale:
	EQLS	2007	Household	Household size?[1]	Scale:
	ESPAD	2007	15 to 16	Who else lives at home? Alone, Father, Mother, Stepfather, Stepmother, Brother, Sister, Other relative, non-relative	
	ESS	2008	Household	What is the size of your household?[1]	Scale:
	EU-SILC	2008	Household	Household size?[1]	Scale:
	HBSC	2002	11, 13, and 15	Now we'd like to ask you about who you live with. Not everyone lives with both their parents. Sometimes people live with just one parent, sometimes they have two homes or two families.	Yes/No: Please fill in column A for your main or your only home. Fill in column B if you have a second home (not including holiday or summer houses). Mother; Father; Stepmother (or father's girlfriend); Stepfather (or mother's boyfriend); Grandmother; Grandfather; I live in a foster home or children's home; Someone or somewhere else: please write it down (open-ended).
	HBSC	2002	11, 13, and 15	Please say how many brothers and sisters live here (including half, step or foster brothers and sisters). Please write in the number or write 0 (zero) if there are none.	Scale: Both homes are included (see other family culture questions) Brothers and Sisters.
	PIRLS	2001	9 to 10	How many people live in your home?	Scale:
	PIRLS	2001	9 to 10	How many children live in your home?	Scale:
	PISA	2000, 2009	15	Who usually lives at <home> with you?	Nominal List:
	PISA	2000	15	How many brothers and sisters do you have?	Scale: Number of brothers and sisters separately.
	TIMSS	2003	10	Including yourself, how many people live in your home?	Scale:
Languages	PIRLS	2001	9 to 10	How often do you speak <language of test> with adults living in your home?	Tick box: Always or almost always; Sometimes; Never.
	PIRLS	2001, 2006	9 to 10	How often do you speak <language of test> at home?	Tick box: Always or almost always; Sometimes; Never.
	PIRLS	2001, 2006	9 to 10	Which languages did you learn to speak when you were little?	Tick box: Test language; Other (list).
	PISA	2000,	15	What language do you speak at home most of the time?	Nominal List:

		2003, 2006, 2009			
	TIMSS	2003	10	How often do you speak English at home?	
	TIMSS	2007	9 to 13	How often do you speak <language> at home?	Likert scale: always; almost always; sometimes; never
Migrant status	EQLS	2007	15+	Were your parents born in this country?	Yes/No:
	EQLS	2007	15+	Where you born in this country?	Yes/No:
	ESPAD	2007	15 to 16	Where were you born?	
	ESS	2008	15+	Are you a citizen of <country>?	Yes/No:
	ESS	2008	15+	How long ago did you come to live in <country>?	Ordinal scale: within the last year; 1-5 years ago; 6-10 years ago; 11-20 years ago; more than 20 years ago; (Don't know)
	ESS	2008	15+	Was your father, mother born in <country>?	Yes/No:
	ESS	2008	15+	Were you born in <country>?	Yes/No:
	EU-SILC	2008	16+	Citizenship?	Yes/No:
	EU-SILC	2008	16+	Migration status (country of birth and parent's country of birth)?	
	PIRLS	2001	9 to 10	How old were you when you came to the test country?	Scale:
	PIRLS	2001, 2006	9 to 10	Was your father born in the test country?	Tick box: Yes; No; don't know.
	PIRLS	2001, 2006	9 to 10	Was your mother born in the test country?	Tick box: Yes; No; don't know.
	PIRLS	2001, 2006	9 to 10	Were you born in the <test country>?	Yes/No:
	PISA	2006, 2009	15	How old were you when you came to the test country?	Scale:
	PISA	2000, 2003, 2009	15	In what country were you and your parents born?	Nominal List: For 2003 this question was expanded to cover countries specific to the country of test.
	PISA	2003	15	Age when arrived in test country	Scale:
	TIMSS	2003	10	How old were you when you came to the test country?	Tick box: Younger than 11 to 55 or older
	TIMSS	2003	10	Was your father (or stepfather or male guardian) born in the test country?	Yes/No:
	TIMSS	2003	10	Was your mother (or stepmother or female guardian) born in the test country?	Yes/No:
	TIMSS	2003	10	Were you born in the <test country>?	Yes/No:
	TIMSS	2007	9 to 13	Country of birth	
	TIMSS	2007	9 to 13	Father's country of birth	
	TIMSS	2007	9 to 13	If not born here, then how old were you when you first arrived?	Ordinal scale: older than 10 years old; 5-10 years old; younger than 5 years old
	TIMSS	2007	9 to 13	Mother's country of birth	
	ESPAD	2007	15 to 16	Father's level of education?	Ordinal scale: ISCED 1-6
	ESPAD	2007	15 to 16	Mother's level of education?	Ordinal scale: ISCED 1-6
Parental education	ESS	2008	15+	Father's level of education?	Tick box: ISCED 1-6
	ESS	2008	15+	Highest level of education?	Ordinal scale: ISCED 1-6
	ESS	2008	15+	Mother's level of education?	Tick box: ISCED 1-6
	ESS	2008	15+	Partner's level of education?	Ordinal scale: ISCED 1-6
	EU-SILC	2008	16+	ISCED level of parents?	Ordinal List: ISCED 1-6
	PISA	2006,9	15	Father's highest ISCED level?	Ordinal scale: ISCED 1-6
	PISA	2006,9	15	Mother's highest ISCED level?	Ordinal scale: ISCED 1-6
	PISA	2000	15	Did your father complete <ISCED 3A>?	Yes/No:
	PISA	2000	15	Did your father complete <ISCED 5A, 5B, 6>?	Yes/No:
	PISA	2000	15	Did your mother complete <ISCED 3A>?	Yes/No:
	PISA	2000	15	Did your mother complete <ISCED 5A, 5B, 6>?	Yes/No:
	TIMSS	2007	13	Father's ISCED level?	Ordinal scale: ISCED 1-6
	TIMSS	2007	13	Mother's ISCED level?	Ordinal scale: ISCED 1-6
	ESS	2008	15+	Do you consider yourself belonging to any particular religion or denomination? Which one?	Yes/No:
	ESS	2008	15+	How religious are you?	Likert scale: Scale of 0-10, with 0=not at all religious and 10=very religious
Religion	ESPAD	2007	15 to 16	Gender?	
	ESS	2008	Household	Gender of all households?[1]	Tick box: Male/Female/(No answer)
Sex	EU-SILC	2008	Household	Gender?[1]	
	PIRLS	2001	9 to 10	Are you a boy or a girl?	Tick box: Male/Female
	PIRLS	2006	9 to 10	Gender?	
	PISA	2000, 2006, 2009	15	Are you <female> or <male>?	Nominal List:
	TIMSS	2003	10	Are you a girl or a boy?	Tick box: Male/Female
	TIMSS	2007	9 to 13	Gender?	Tick box: Male/Female

Education and schooling

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Attainment and achievement	ESPAD	2007	15 to 16	Your average grade?	Ordinal scale: A to F
	ESPAD	2007	15 to 16	How good at school work?	Ordinal scale:
	PIRLS	2001 and 2006	9 to 10	How well do you read? Tell how much you agree with each of these statements.	Likert scale (Agree a lot, Agree a little, Disagree a little, Disagree a lot): Reading is very easy for me; I do not read as well as other students in my class; When I am reading by myself, I understand almost everything I read; Reading aloud is very hard for me.
	PISA	2006	15	How easy will it be to perform the following tasks (recognise science question in health article, explain earthquakes, explain antibiotics, identify science related to disposing garbage, recognise environment science, understand scientific information on food labels, science related to life on Mars, identify main reason for acid rain)	Likert scale: could do this easily; could do this with a bit of effort; would struggle to do this on my own; could not do this.
	PISA	2000	15	In your last school report, what <grade> did you receive in the following subjects?	Ordinal scale: A to F
	PISA	2000	15	In your last school report, how did your <grade> compare with the <pass grade> in each subject area?	Likert scale: Above, on, below for each literacy type
	PISA	2003	15	Have you ever repeated a grade?	Ordinal list of grades: no, once, more than once.
	PISA	2009	15	Have you ever repeated a grade in <ISCED 1>, <ISCED2> or <ISCED3>?	
	PISA	2000	15	What <grade> are you in?	Number:
Attendance	ESPAD	2007	15 to 16	During last 30 days how many times have you missed a lesson?	Scale:
	PIRLS	2006	9 to 10	Do you go to school?	Yes/No:
	PISA	2000	15	In the last full week you were in school, how many <class periods> did you spend in:	Scale:
	PISA	2000	15	On average, about how many students are in your:	Scale:
	TIMSS	2007	9 and 13	Do you go to school?	Yes/No:
Early years	PISA	2003, 2009	15	Did you attend pre-school <ISCED 0>	Yes/No:
	PISA	2003	15	Age when started year 1 of primary	Yes/No:
Extra curricula	PISA	2000	15	During the last three years, have you attended any of these special courses outside of your school to improve your results?	Ordinal List (none to 3 or more courses): in <test language>; courses in other subjects; <Extension> or additional courses; <Remedial> courses in <test language>; <Remedial> courses in other subjects; Training to improve your study skills; <Private tutoring>.
	PISA	2006, 2009	15	What type of out of school lessons do you attend? (teacher at school, other teacher, small group with teacher at school, small group with other teacher, large group with teacher at school, large group with other teacher)	Yes/No:
	TIMSS	2003	10	During this school year, how often have you had tutoring or extra lessons in mathematics that are not part of your regular class?	Likert scale: Every or almost every day; Once or twice a week; Sometimes; Never or almost never. Same question is asked about Science.
General education	EQLS	2007	15+	How old when you completed full time education?	Scale:
	EQLS	2007	15+	Highest level of education?	Ordinal scale: ISCED 1-6
	ESS	2008	15+	In which field is your highest level of education?	Tick box: List of subject areas
	ESS	2008	15+	How many years of education have you completed?	Scale:
	EU-SILC	2008	Household	Education at pre-school and compulsory school[1]	Yes/No:
	EU-SILC	2008	16+	ISCED level	Ordinal List: ISCED 1-6
	PISA	2006	15	Do you take any of the courses listed below? (compulsory/optional biology/physics/chemistry	Yes/No:
Homework	PIRLS	2001	9 to 10	How often does your teacher give you reading to do for homework (for any subject)?	Tick box: I never have reading to do for homework; Less than once a week; 1 or 2 times a week; 3 or 4 times a week; Every day.
	PIRLS	2006	9 to 10	How often do you get reading homework?	Ordinal scale: Never; less than once a week; 1-2 times a week; 3-4 times a week; every day.
	PIRLS	2006	9 to 10	Who helps you with reading homework?	Tick box: Never have homework; never need help; parents/grandparents; brother/sister; teacher/tutor; someone else; there is nobody to help.

	PISA	2000	15	Please indicate how often each of these applies to you.	Likert scale (Never to always): My homework is counted as part of my <marks>; I am given interesting homework; My teachers make useful comments on my homework; I finish my homework during the school day; My teachers grade my homework; I do my homework while watching television; I complete my homework on time.
	PISA	2000	15	On average, how much time do you spend each week on homework and study in these subject areas?	Likert scale (hours per week): all literacy types.
	TIMSS	2003	10	How often does your teacher give you homework in mathematics?	Likert scale (Every day, 3 or 4 times a week, or 2 times a week, Less than once a week, never): the same question is asked for science.
	TIMSS	2003	10	When your teacher gives you mathematics homework, about how long does it take you to complete this homework?	Likert scale (Fewer than 5 minutes, 5-30 minutes, 3-60 minutes, 6-90 minutes, More than 90 minutes): Science also.
Teaching and learning	ICCS	2009	13 to 14	Do your teachers encourage students to make up their own mind?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	ICCS	2009	13 to 14	Do your teachers encourage students to express their opinion?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	ICCS	2009	13 to 14	Do other students bring up current political situations for discussion in class?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	ICCS	2009	13 to 14	Do other students express opinions in class even if they are different from most other students?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	ICCS	2009	13 to 14	Do teachers encourage students to discuss issues with people with different opinions?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	ICCS	2009	13 to 14	Do teachers present several sides of an issue when explaining them in class?[2]	Likert scale: Strongly disagree; disagree; agree; strongly disagree.
	PIRLS	2001	9 to 10	In school, how often do these things happen?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost never): my teacher reads aloud to the class; I read aloud to the whole class; I read aloud to a small group of students in my class; I read silently on my own; I read along silently while other students read aloud; I read books that I choose myself.
	PIRLS	2001	9 to 10	After you have read something in class, how often do you do these things?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost never): I answer questions in a workbook or on a worksheet about what I have read ; I write something about what I have read (for example, a summary, a story, or how I felt about what I read) ; I answer questions aloud that my teacher asks about what I have read ; I talk with other students about what I have read ; I draw pictures or do an art project about what I have read ; I act in a play or drama about what I have read ; I do a group project with other students in the class about what I have read ; I take a written quiz or test about what I have read.
	PIRLS	2006	9 to 10	In school, how often: teacher reads aloud; you read aloud; read silently; read books that you choose?	Ordinal scale: every day/almost every day; once/twice a week; once/twice a month; never/almost never
	PIRLS	2006	9 to 10	After reading, how often do you: answer question in workbook, write something about the book; answer questions aloud; talk to other students about the book	Likert scale: every day/almost every day; once/twice a week; once/twice a month; never/almost never
	PISA	2000	15	How often do these things happen in your <test language> lessons?	Ordinal List (never / some / most to every lesson): The teacher has to wait a long time for students to <quieten down>; 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; 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; The teacher helps students with their learning; The teacher checks students' homework; 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; Students have to learn a lot; There is noise and disorder; At the start of class, more than five minutes are spent doing nothing.
	PISA	2000	15	At your school, about how often do you use: school library?; computers?; calculators?; Internet?; <science> laboratories?	Likert scale: Never or hardly ever ; several times a week.

Health and risks

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Body and image	HBSC	2002	11, 13, and 15	How much do you weigh without clothes?	Scale: Kilograms (kg) or Pounds (lbs).
	HBSC	2002	11, 13, and 15	How tall are you without shoes?	Scale: in centimeters or feet and inches.
	HBSC	2002	11, 13, and 15	At present are you on a diet or doing something else to lose weight?	Likert scale: No, my weight is fine; No, but I should lose some weight; No, because I need to put on weight; Yes.
General health	EQLS	2007	15+	How satisfied with your own health?	Ordinal scale: 1-10 (1=very dissatisfied and 10=very satisfied)
	EU-SILC	2008	16+	General health	Ordinal List:
	EU-SILC	2008	16+	Limitation of activities due to health?	Yes/No:
	HBSC	2002	11, 13, and 15	In the last 6 months: how often have you had the following…?	Likert scale (About every day, More than once a week, About every week, About every month, Rarely or never): Headache; Stomach-ache; Back ache; Feeling low; Irritability or bad temper; Feeling nervous ; Difficulties in getting to sleep; Feeling dizzy.
Mental health, disability and chronic illness	EQLS	2007	15+	Any chronic physical or mental health problem, illness or disability?	Yes/No:
	ESPAD	2007	15 to 16	Last week, how often? Lost appetite; difficulty concentrating; felt depressed; pressurised; sad; couldn't do work	Ordinal scale:
	ESPAD	2007	15 to 16	How many times self harm / attempted suicide?	Ordinal scale:
	ESS	2008	15+	Do you have long-standing illness or disability?	Yes/No:
	EU-SILC	2008	16+	Suffer from any chronic illness?	Yes/No:
Nutrition	HBSC	2002	11, 13, and 15	How often do you usually have breakfast (more than a glass of milk or fruit juice*[1])?	Likert scale (none through to every day): Instead of fruit juice or milk (or country specific examples can be given). The question is repeated for weekends (only three options).
	HBSC	2002	11, 13, and 15	How often do you usually have lunch (midday meal) (more than a drink or a snack)?	Likert scale (none through to every day) :The question is repeated for weekends (only three options).
	HBSC	2002	11, 13, and 15	How often do you usually have supper (evening meal) (more than a drink or a snack)?	Likert scale (none through to every day):The question is repeated for weekends (only three options).
	HBSC	2002	11, 13, and 15	How many times a week do you usually eat or drink…?	Likert scale (Never, less than once a week, once a week, 2-4 days a week, 5-6 days a week, once a day, every day, and Every day, more than once): Fruits; Vegetables; Sweets (candy or chocolate); Coke or other soft drinks that contain sugar; Any alcoholic drink.
Personal care and injury	HBSC	2002	11, 13, and 15	How often do you brush your teeth?	Likert scale: More than once a day, Once a day, At least once a week but not daily, Less than once a week, Never.
	HBSC	2002	11, 13, and 15	During the past 12 months, how many times were you injured and had to be treated by a doctor or nurse?	Likert scale: I was not injured in the past 12 months, 1 time, 2 times, 3 times, 4 times or more.
	HBSC	2002	11, 13, and 15	Over a typical or usual week, on how many days are you physically active for a total of at least 60 minutes per day?	Likert scale: Also asks for recent week data.
Risk behaviour: general	ESPAD	2007	15 to 16	How much money spent over last 30 days on? Tobacco, alcohol, cannabis	Ordinal scale: Euros 0; 1-3; 4-6; 7-15; 16-30; 31-70 71 and more
	ESPAD	2007	15 to 16	How many friends do? Smoke; drink; get drunk; cannabis; ecstasy; inhalants	Ordinal scale: None; a few; some; most; all
	ESPAD	2007	15 to 16	Do any siblings do? Smoke; drink; get drunk; cannabis; ecstasy; inhalants	Yes/No:
	HBSC	2002	11, 13, and 15	At what age did you first do the following things?	Scale: Drink alcohol (more than a small amount); Get drunk; Smoke a cigarette (more than a puff).
Risk behaviour: drinking	ESPAD	2007	15 to 16	How difficult to get alcohol?	Likert scale: impossible; very difficult; fairly difficult; fairly easy; very east; (Don't know)
	ESPAD	2007	15 to 16	How often do you drink alcohol?	Ordinal scale:
	ESPAD	2007	15 to 16	How much did you drink last time you drank?	Ordinal scale:
	ESPAD	2007	15 to 16	How drunk were you the last time you drank?	Ordinal scale: 1-10 (10=heavily intoxicated e.g. not remembering what happened)
	ESPAD	2007	15 to 16	How often did you buy alcohol in the last 30 days?	Ordinal scale:
	ESPAD	2007	15 to 16	Over the last 30 days how many times have you drank more than 5 drinks?	Ordinal scale:
	ESPAD	2007	15 to 16	How many times have you been drunk in your lifetime / last year / last 30 days?	Ordinal scale:

	ESPAD	2007	15 to 16	How likely that the following will happen if drinking? Relaxed; Police trouble; Harm health; Happy; Forget problems; Not be able to stop drinking; Hangover; Feel friendly; Do something regretful; Have fun; Feel sick.	Likert scale: very likely; likely; unsure; unlikely; very unlikely
	ESPAD	2007	15 to 16	Because of alcohol, how often? Fight; injury; problem with parents; problem with friends; poorly at school; trouble with police; hospitalised; had sex	Ordinal scale:
	ESPAD	2007	15 to 16	Where did you have your last drink?	Tick box: I never drink; at school; at home; someone else's home; out on the street; bar/pub; disco; restaurant; other
	HBSC	2002	11, 13, and 15	At present, how often do you drink anything alcoholic, such as beer, wine or spirits like	Likert scale: Beer; Wine; Spirits; (National drinks categories can be added as appropriate).
	HBSC	2002	11, 13, and 15	Have you ever had so much alcohol that you were really drunk?	Likert scale (No, never, Yes, once. Yes, 2-3 times, Yes, 4-10 times, Yes, more than 10 times).
Risk behaviour: drugs	ESPAD	2007	15 to 16	How often taken tranquilisers or sedatives?	Ordinal scale:
	ESPAD	2007	15 to 16	How difficult is it to get cannabis?	Likert scale: impossible; very difficult; fairly difficult; fairly easy; very east; (Don't know)
	ESPAD	2007	15 to 16	How many times have you used cannabis? Lifetime/last year/last 30 days	Ordinal scale:
	ESPAD	2007	15 to 16	When was the first time you used cannabis?	Ordinal scale:
	ESPAD	2007	15 to 16	Have you ever refused use of cannabis?	Yes/No:
	ESPAD	2007	15 to 16	How difficult to get amphetamines/sedatives/ecstasy/inhalants?	Likert scale: impossible; very difficult; fairly difficult; fairly easy; very east; (Don't know)
	ESPAD	2007	15 to 16	How many times have you used ecstasy? Lifetime/last year/last 30 days	Ordinal scale:
	ESPAD	2007	15 to 16	How many times have you used inhalants? Lifetime/last year/last 30 days	Ordinal scale:
	ESPAD	2007	15 to 16	How many times have you used the following drugs in your lifetime? Sedatives; amphetamines; LSD; crack; cocaine; relevin; heroin; mushrooms; ghb; steroids; others	Ordinal scale:
	ESPAD	2007	15 to 16	When was the first time you used drugs?	Ordinal scale:
	ESPAD	2007	15 to 16	Because of drugs, how many times? Fight; injury; problem with parents; problem with friends; poorly at school; trouble with police; hospitalised; had sex	Ordinal scale:
	ESPAD	2007	15 to 16	If used cannabis, would you admit to it in questionnaire?	Ordinal scale:
	HBSC	2002	15	Have you ever taken cannabis…?	Likert scale (Never, Once or twice, 3 to 5 times, 6 to 9 times, 10 to 19 times, 20 to 39 times, 40 times or more - both in your life; Last 12 months).
Risk behaviour: sex	HBSC	2002	15	Have you ever had sexual intercourse (sometimes this is called "making love," "having sex," or "going all the way" or other appropriate colloquial terms)?	Yes/No:
	HBSC	2002	15	How old were you when you had sexual intercourse for the first time?	Likert scale (I have never had sexual intercourse, 11 years old or younger, 12 years old, 13 years old, 14 years old, 15 years old, 16 years old).
	HBSC	2002	15	The last time you had sexual intercourse, did you or your partner use a condom?	Yes/No: Includes a 'never had sexual intercourse' option.
	HBSC	2002	15	The last time you had sexual intercourse, what method(s) did you or your partner use to prevent pregnancy?	Yes/No: Includes an initial question to route those not using any method and those not having had sex. Birth control pills; Condoms; Spermicidal spray or foam; Withdrawal; (National choice option here); Some other method; Not sure.
Risk behaviour: Smoking	ESPAD	2007	15 to 16	How difficult is it to get cigarettes?	Likert scale: impossible; very difficult; fairly difficult; fairly easy; very east; (Don't know)
	ESPAD	2007	15 to 16	How many times have you smoked in your lifetime / last year / last 30 days?	Ordinal scale:
	ESPAD	2007	15 to 16	How frequently have you smoked in last 30 days?	Ordinal scale:
	ESPAD	2007	15 to 16	When was the first time you did the following? Smoke; Smoke regularly	Ordinal scale:
	ESPAD	2007	15 to 16	Would father/mother allow smoking?	Likert scale:
	HBSC	2002	11, 13, and 15	Have you ever smoked tobacco?	Yes/No:
	HBSC	2002	11, 13, and 15	How often do you smoke tobacco at present?	Likert scale (Every day, At least once a week, but not every day, Less than once a week, I do not smoke).

Income and deprivation

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Costs	EQLS	2007	15+	Is total housing cost a financial burden?	Likert scale: a heavy burden; somewhat of a burden; not burden at all; (Don't know).
	EU-SILC	2008	16+	Financial burden of total housing cost?	Scale:
	EU-SILC	2008	16+	Interest payment?	Scale:
Debt	EQLS	2007	15+	Has your household been in arrears in the last 12 months on? Rent or mortgage / utility bills	Yes/No:
	EU-SILC	2008	16+	Arrears on loans?	Yes/No:
	EU-SILC	2008	16+	Arrears on mortgage payment?	Yes/No:
	EU-SILC	2008	16+	Arrears on utility bills?	Yes/No:
Deprivation (cultural)	PIRLS	2001, 2006	9 to 10	About how many books are there in your home?	Tick box (0-10; 11 to 25; 26-100; 101-200; 200 or more).
	PISA	2000, 2003, 2006, 2009	15	How many books are there in your home?	Ordinal List (grouped for 0-10 to 500+).
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	How many books are there in your home?	Scale:
Deprivation (educational)	EU-SILC	2008	16+	Do you have a computer?	Yes/No:
	HBSC	2002	11, 13, and 15	How many computers does your family own?	Ordinal List (None, one, two, more than two).
	TIMSS	2007	9 and 13	Do you have any of these things at home? Calculator, computer, study desk, dictionary	Yes/No:
Deprivation (food)	EQLS	2007	15+	Has household ever run out of money to pay for food in the last 12 months?	Yes/No:
	EU-SILC	2008	16+	Able to afford protein based meal?	Yes/No:
	HBSC	2002	11, 13, and 15	Some young people go to school or to bed hungry because there is not enough food at home. How often does this happen to you?	Ordinal List (Always, Often, Sometimes, Never).
Deprivation (General)	ESS	2008	15+	In next 12 months how likely that you won't have enough money to cover household necessities?	Likert scale: Not at all likely; not very likely; likely; very likely; (Don't know)
	ESS	2008	15+	How do you feel about your current households income?	Likert scale: Can live comfortably; coping; finding it difficult; finding it very difficult; (Don't know)
	EU-SILC	2008	16+	Able to pay for one week of annual holiday?	Yes/No:
	EU-SILC	2008	16+	Do you have a car?	Yes/No:
	EU-SILC	2008	16+	Do you have a colour TV?	Yes/No:
	EU-SILC	2008	16+	Do you have a telephone?	Yes/No:
	EU-SILC	2008	16+	Do you have a washing machine?	Yes/No:
	EQLS	2007	15+	Can you afford if you want? (keep house adequately warm/week's annual holiday/replace damaged furniture/protein meal/new clothes/having friends over once a month)	Yes/No:
	HBSC	2002	11, 13, and 15	Does your family own a car, van or truck?	Ordinal List (No, Yes one, more than one).
	HBSC	2002	11, 13, and 15	During the past 12 months, how many times did you travel away on holiday (vacation) with your family?	Ordinal List (Not at all, Once, Twice, More than twice).
	PIRLS	2001, 2006	9 to 10	Do you have any of these things in your home?	Yes/No: Computer (do not include Nintendo®, Gameboy®, or other TV/video game computers) ; Study desk/table for your use ; Books of your very own (do not count your school books) ; Daily newspaper ; <country-specific indicator of wealth> (8 options).
	PISA	2000, 2003, 2006, 2009	15	In your home, do you have:	Yes/No: a dishwasher; a room of your own; educational software; a link to the Internet; dictionary; a quiet place to study; a desk for study; text books; classic literature (e.g., <Shakespeare>); books of poetry; works of art (e.g., paintings); For 2003 onwards 3 country specific items could be included.
	PISA	2000, 2003, 2006, 2009	15	How many of these do you have at your home?	Ordinal list (none to 3 or more): <Cellular> phone ; Television; Calculator; Computer ; Musical instrument (e.g., piano, violin) ; Motor car ; Bathroom.
	TIMSS	2003	10	Do you have any of these items in your home?	Yes/No: Calculator; Computer; Study desk/table for your use; Dictionary; Encyclopedia (as a book or CD); PlayStation®, Game Cube®, XBox®, or other TV/video game; DVD Player; Three or

					more cars, small trucks, or sport utility vehicles; Country specific (8 options).
Employment security	ESS	2008	15+	Have you ever been unemployed and seeking employment for period of 3 months or more?	Yes/No:
	ESS	2008	15+	If not in paid work, then when was the last year you had paid work?	Scale:
	ESS	2008	15+	In the next 12 months how likely are you to be unemployed?	Likert scale: Not at all likely; not very likely; likely; very likely; (never worked/no longer working and not looking for work); (Don't know)
	ESS	2008	15+	In the next 12 months how likely that you will less time in paid work than you would like?	Likert scale: Not at all likely; not very likely; likely; very likely; (never worked/no longer working and not looking for work); (Don't know)
Housing conditions	EU-SILC	2008	16+	Ability to warm house adequately?	Yes/No:
	EU-SILC	2008	16+	Bath or shower in dwelling?	Yes/No:
	HBSC	2002	11, 13, and 15	Do you have your own bedroom for yourself?	Yes/No:
Income and benefits	EQLS	2007	15+	Has anyone in your household received any of the following in the last 12 months (salary/personal profit/pension/child benefit/unemployment benefit/other income)?	Yes/No:
	EQLS	2007	15+	What is the household's total disposable income?	Scale:
	EQLS	2007	15+	Which income source is the largest? (salary/personal profit/pension/child benefit/unemployment benefit/other income)	Tick box:
	ESPAD	2007	15 to 16	How much money spent on personal needs with parent's control per week?	Scale:
	ESS	2008	15+	Households total income?	Tick box: ten options.
	ESS	2008	15+	What is the main source of income in your household? Wage / income from self-employment / income from farming / pensions / benefit / other social grants / income from investment / other sources	Tick box:
	EU-SILC	2008	16+	Income from property; investment?	Scale:
	EU-SILC	2008	16+	Income received by people aged under 16?	Scale:
	EU-SILC	2008	16+	Minimum monthly income required to make ends meet?	Scale:
	EU-SILC	2008	16+	Total disposable income before taxes and transfers?	Scale:
	EU-SILC	2008	16+	Total disposable income?	Scale:
	EU-SILC	2008	16+	Total household income?	Scale:
	EU-SILC	2008	16+	Education related allowance	Scale:
	EU-SILC	2008	16+	Household disability benefit?	Scale:
Parental employment	EQLS	2007	15+	Hours of work?	Scale:
	EQLS	2007	15+	Type of employment (permanent, part-time, apprenticeship, without written contract)?	Tick box:
	EQLS	2007	15+	Current/last employment type?	Tick box:
	EQLS	2007	Household	Employment status of each household?[1]	Yes/No:
	EQLS	2007	15+	If not in paid work, then ever had a paid job?	Yes/No:
	ESS	2008	15+	Employment status? Paid work / education / unemployed and looking for work / unemployed but not looking for a job / permanently sick or disabled / military service / housework / other	Tick box:
	ESS	2008	15+	Duration of employment? Unlimited / Limited	Tick box:
	ESS	2008	15+	What are you contracted hours per week (excluding overtime)	Scale:
	ESS	2008	15+	Partner's employment status?	Tick box:
	EU-SILC	2008	16+	Employment status of adults in household?	
	EU-SILC	2008	16+	Total number of hours worked by parents?	Scale:
	HBSC	2002	11, 13, and 15	Does your <see sub question> have a job?	Nominal List (Yes, No, Don't Know, don't have or don't see <person>, Father or Mother).
	HBSC	2002	11, 13, and 15	If NO, why does your <person> not have a job?	Nominal List (He is sick, or retired, or a student, He is looking for a job, He takes care of others, or is full-time in the home, I don't know Father; Mother).
	HBSC	2002	11, 13, and 15	If YES, please say in what place he/she works	Open ended: for both Father and Mother.
	HBSC	2002	11, 13, and 15	Please write down exactly what job he does there	Open ended: for both Father and Mother.
	PISA	2006, 2009	15	Father's employment status?	
	PISA	2000, 2009	15	What does your father do in his main job? (e.g., <Teaches high school students, builds houses, manages a sales team>)	Open ended:
	PISA	2000, 2009	15	What does your mother do in her main job? (e.g., <Teaches high school students, cares for patients, manages a sales team>)	Open ended:
	PISA	2000	15	What is your father currently doing?	Open ended:

	PISA	2000	15	What is your father's main job? (e.g., <School teacher, carpenter, sales manager>)	Open ended:
	PISA	2000	15	What is your mother currently doing?	Open ended:
	PISA	2000	15	What is your mother's main job (or most recent)? (e.g., <School teacher, nurse, sales manager>)	Open ended:
	PISA	2006, 2009	15	Mother's employment status?	
Support	EQLS	2007	15+	Has your household received regular financial help in the last year from relatives or friends?	Yes/No:
	EQLS	2007	15+	In the past year, has your household given food or money to a relative not living in the house?	Yes/No:
Taxes	EU-SILC	2008	16+	Tax adjustment?	Scale:
	EU-SILC	2008	16+	Tax on income and social contributions?	Scale:
	EU-SILC	2008	16+	Taxes on wealth?	Scale:

Housing and environment

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Crime	ESS	2008	15+	Has a member of your household been the victim of burglary in the last 5 years?	Yes/No:
	ESS	2008	15+	Do you feel safe walking alone in this area?	Yes/No:
	EU-SILC	2008	16+	Crime violence in the area?	Yes/No:
	HBSC	2002	11, 13, and 15	Generally speaking I feel safe in the area I live	Likert scale (Always, most of the time, Sometimes, rarely or Never).
Environment	EQLS	2007	15+	Complains about local area? (noise / air pollution / lack of recreational area / water quality / crime / litter and rubbish)	Likert scale:
	EU-SILC	2008	16+	Noise from neighbours or street?	Yes/No:
	EU-SILC	2008	16+	Pollution, crime or other environmental problem?	Yes/No:
Housing conditions	EQLS	2007	15+	How many rooms in accommodation excluding kitchen, bathrooms, halls, storerooms, and rooms for business?	Scale:
	EQLS	2007	15+	Any of the following problems with accommodation? (shortage of space/damage/lack of indoor toilet/lack of bath shower/lack of place to sit outside)	Yes/No:
	EU-SILC	2008	16+	Damage to house?	Yes/No:
Housing security	EQLS	2007	15+	How likely that you will have to leave the accommodation in the 6 months because you can't afford it?	Likert scale: Very likely; quite likely; quite unlikely; very unlikely; (Don’t know)
Locality	EQLS	2007	15+	Does your area have? (food store / post office / banking/ cinema / public transport / recycling facility)	Yes/No:
	EQLS	2007	15+	The area you live in is? (open countryside / village / medium town / city or city suburb)	Tick box:
	ESS	2008	15+	What best describes where you live? Big city, suburb of big city, town, village, countryside	Tick box:
Main home	HBSC	2002	11, 13, and 15	Do you stay here…… (see homes / family culture)	Likert scale (All the time, Most of the time, Half the time). For the second home (Half the time, Regularly but less than half the time, At weekends, Sometimes, Hardly ever).
Tenure	EQLS	2007	15+	Accommodation ownership status? (own without mortgage / own with mortgage / renting from private landlord / renting from social housing / rent free / other)	Tick box:
	EU-SILC	2008	16+	Residential status; tenure status?	Ordinal scale:
	EU-SILC	2008	16+	Dwelling Type?	Ordinal scale:
	EU-SILC	2008	16+	Tenure status?	Yes/No:

Civic participation and time use

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Civic participation (political)	EQLS	2007	15+	Did you vote in the last national election?	Yes/No:
	ESS	2008	15+	Do you vote?	Yes/No:
	ESS	2008	15+	In which ways have you stopped wrong things? Contacted politician, worked in political group, signed a petition, demonstration, boycott product	Yes/No:
	ESS	2008	15+	Are you a member of a political party?	Yes/No:
	ESS	2008	15+	Interest in politics	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
	ESS	2008	15+	Trust in public institutions?	Likert scale: 0-10 (10=complete trust)
	ICCS	2009	13 to 14	How interested are you in political issues?[2]	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
	ICCS	2009	13 to 14	How interested are you in international politics?[2]	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
	ICCS	2009	13 to 14	When you are an adult will you help a candidate in an election campaign?[2]	Likert scale: I will certainly not do this; I will probably not do this; I will probably do this; I will certainly do this
	ICCS	2009	13 to 14	When an adult will you join a political party?[2]	Likert scale: I will certainly not do this; I will probably not do this; I will probably do this; I will certainly do this
	ICCS	2009	13 to 14	When an adult will you join a trade union?[2]	Likert scale: I will certainly not do this; I will probably not do this; I will probably do this; I will certainly do this
	ICCS	2009	13 to 14	When an adult will you stand in any elections?[2]	Likert scale: I will certainly not do this; I will probably not do this; I will probably do this; I will certainly do thisv
Civic participation (social)	CIVED	1999	14	Have you participated in the following organisations?	Yes/No: A student council/student government [class or school parliament]; A youth organisation affiliated with a political party or union; A group which prepares a school newspaper; An environmental organisation; A U. N. or UNESCO Club; A student exchange or school partnership program; A human rights organisation; A group conducting [voluntary] activities to help the community; A charity collecting money for a social cause Boy or Girl Scouts [Guides]; A cultural association [organisation] based on ethnicity; A computer club; An art, music or drama organisation; A sports organisation or team; An organisation sponsored by a religious group.
	CIVED	1999	14	Think about all the organisations listed above. How often do you attend meetings or activities for any or all of these organisations?	Likert scale: every day through to never
	EQLS	2007	15+	How often do the following activity? (Educational care for children / cooking / care for elderly / voluntary work)	Ordinal scale:
	EQLS	2007	15+	How many hours per week on? (Educational care for children / cooking / care for elderly / voluntary work)	Ordinal scale:
	ESS	2008	15+	In the last month have you done any paid/voluntary work?	Tick box:
	ICCS	2009	13 to 14	How interested are you in social issues?[2]	Likert scale: Not interested at all; Not very interested; Quite interested; Very interested
Education time use	HBSC	2002	11, 13, and 15	About how many hours a day do you usually spend doing school homework out of school hours?	Likert scale (none through to 7 or more hours): Split for weekday and weekend.
	PIRLS	2001	9 to 10	On days when you have reading to do for homework (for any subject), how much time do you spend on this reading?	Tick box: I never have reading to do for homework; Half hour or less; Between a half hour and one hour; One hour or more.
	PIRLS	2006	9 to 10	When you have reading homework, how much time do you spend?	Likert scale:
	PIRLS	2001, 2006	9 to 10	How often do you do these things outside of school?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost never): I read aloud to someone at home; I listen to someone at home read aloud to me; I talk with my friends about what I am reading ; I talk with my family about what I am reading ; I read for fun outside of school ; I read to find out about things I want to learn; I watch television or videos outside of school.
	PIRLS	2001, 2006	9 to 10	How often do you read these things outside of school?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost

					never): I read comic books ; I read stories or novels; I read books that explain things (You might read about your favorite athlete, about animals you like, or a place you visited.); I read magazines; I read newspapers ; I read directions or instructions (You might read them to put a toy together, to learn how to play a game, or to do something else.); <I read subtitles on the television screen>; <country-specific>.
	PISA	2006	15	How often do you do these things? (TV science, books science, website science, radio science, read science, attend science)	Likert scale:
	PISA	2000	15	How often do the following people work with you on your <schoolwork>? List of relatives.	Ordinal scale:
	PISA	2006	15	On average, how much time do you spend each week on homework and study in these subject areas?	Likert scale: hours per week for all literacy types.
Friendships	CIVED	1999	14	How often do you spend time [directly] after school talking [hanging out] with your friends?	Likert scale (every day through to never).
	CIVED	1999	14	How often do you spend time during the evening [after dinner or after—] outside your home with your friends?	Likert scale (every day through to never).
	HBSC	2002	11, 13, and 15	How many days a week do you usually spend time with friends right after school?	Scale: None through until 5
	HBSC	2002	11, 13, and 15	How many evenings per week do you usually spend out with your friends?	Scale (None through until 7).
	HBSC	2002	11, 13, and 15	How often do you talk to your friend(s) on the phone or send them text or email messages?	Likert scale (Rarely or never, 1 or 2 days a week, 3 or 4 days a week, 5 or 6 days a week, Every day).
Personal time use	CIVED	1999	14	How much time do you spend watching television or videos on school days?	Likert scale (no time to more than 5 hours).
	ESPAD	2007	15 to 16	How often do you do the following? Computer games; Sports; Read books; Socialise; Other hobbies; Shopping; Browse; Slot machines.	Likert scale:
	ESPAD	2007	15 to 16	What housework you usually do?	Tick box: Shopping; care; cook; clean; laundry; dishes; gardening; rubbish disposal; Don't do any work
	ESPAD	2007	15 to 16	How much TV per week?	Ordinal scale:
	ESS	2008	15+	Hours per week? Watching TV, listening to radio, reading newspapers, browsing www,	Ordinal scale:
	ESS	2008	15+	Compared to other people of your age, how often do you take part in social activities	Likert scale: Much less than most; Less than most; About the same; More than most; Much more than most; (Don't know).
	HBSC	2002	11, 13, and 15	About how many hours a day do you usually watch television (including videos) in your free time?	Likert scale (none through to 7 or more hours): Split for weekday and weekend (two questions).
	HBSC	2002	11, 13, and 15	About how many hours a day do you usually use a computer (for playing games, emailing, chatting or surfing the internet) in your free time?	Likert scale (none through to 7 or more hours): Split for weekday and weekend (two questions).
	PIRLS	2001	9 to 10	About how much time do you spend watching television or videos outside of school on a normal school day?	Tick box: No time; Up to 1 hour; From 1 hour up to 3 hours; From 3 hours up to 5 hours; 5 hours or more.
	PIRLS	2001	9 to 10	How often do you use a computer in each of these places?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost never): I use a computer at home ; I use a computer at school ; I use a computer at some other place.
	PIRLS	2001	9 to 10	How often do you do these things with a computer?	Likert scale (Every day, Once or twice a week, Once or twice a month, Never or almost never): I play computer games; I use the computer to write reports or stories; I use the computer to look up information (Internet, CD-ROM); I send and read e-mails.
	PIRLS	2006	9 to 10	How often do you? Listen to music, talk with friends, play sports	Likert scale:
	PIRLS	2001 and 2006	9 to 10	How often do you borrow books from your school or local library to read for fun?	Tick box: At least once a week; Once or twice a month; Never or almost never.
	PIRLS	2001, 2006	9 to 10	How often do you use the internet to do: school information; sports; music; other activities; chat/email?	Likert scale: every day/almost every day; once/twice a week; once/twice a month; never/almost never
	PIRLS	2001, 2006	910	Do you ever use a computer? (Do not include Nintendo, GameBoy, or other TV/video game computers.)	Yes/No:
	PISA	2000	15	Each day, about how much time do you usually spend reading for enjoyment?	Yes/No: I do not read for enjoyment; 30 minutes or less each day; More than 30 minutes to less than 60 minutes each day; 1 to

					2 hours each day; More than 2 hours each day.
	PISA	2000	15	How often do you read these materials because you want to?	Likert scale: Never or hardly ever / several times a week Magazines; Comic ; books; Fictions (novels, narratives, stories); Non-fiction books; Emails and Web pages; Newspapers.
	PISA	2000	15	During the past year, how often have you participated in these activities: Attended a popular music concert.; Attended an opera, ballet or classical symphony concert.; Watched live theatre; Attended sporting events.	Likert scale: Gone to the <pictures>; Visited a museum or art gallery.
	PISA	2000	15	How often do you borrow books to read for pleasure from a public or school library?	Likert scale (Never or hardly ever / several times a week).
	TIMSS	2007	9 and 13	How often do you? Listen to music, talk with friends, play sports	Likert scale: every day/almost every day; once/twice a week; once/twice a month; never/almost never
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	On a normal school day, how much time do you spend before or after school doing each of these things?	Likert scale: I watch television and videos; play computer games; I play or talk with friends; do jobs or chores at home; play sports; read a book for enjoyment; Use the internet; do homework.
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	Do you ever use a computer? (Do not include PlayStation®, GameCube®, XBox®, or other TV/video game system).	Yes/No:
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	Where do you use a computer?	Yes/No: At home; At school; At a library; Friends home; Internet café; Elsewhere.
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	How often do you do these things with a computer?	Likert scale (Every day, At least once a week, Once or twice a month, A few times a year, Never): I look up ideas and information for mathematics; I look up ideas and information for science; I write reports for school.
Religion time use	EQLS	2007	15+	How often do you attend religious services (apart from wedding, funerals and other important religious events)?	Ordinal scale:

Relationships

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Family and social relationships	EQLS	2007	15+	How often have face-to-face contact with relatives leaving outside the house? (children / parents / siblings / friends or neighbours)	Ordinal scale:
	EQLS	2007	15+	How often have contacts by phone/email/post ? (children / parents / siblings / friends or neighbours)	Ordinal scale:
	ESS	2008	15+	How often to you meet friends/family socially?	Ordinal scale:
	ESS	2008	15+	Is there anyone with whom you can discuss personal matters?	Yes/No:
	HBSC	2002	11, 13, and 15	How easy is it for you to talk to the following persons about things that really bother you?	Likert scale (very easy, easy, difficult, very difficult, don't have or see this person): Father; stepfather (or mother's boyfriend); Mother; Stepmother (or father's girlfriend); Elder brother (s); Elder sister (s); Best friend; Friends of the same sex; Friends of the opposite sex.
	HBSC	2002	11, 13, and 15	At present, how many close male and female friends do you have?	Likert scale (None, one, two three or more): Male and Female.
	EQLS	2007	15+	Who can you get support from? (partner or spouse / work colleague / friend / neighbour / someone else / nobody)	Yes/No:
	ESPAD	2007	15 to 16	How often do the following happen? parents set rule; parent know where I am in the evenings; emotional support from family; borrow money; get money as a gift; emotional support from friend	Ordinal scale:
	PISA	2000	15	In general, how often do your parents:	Likert scale (???): Discuss political or social issues with you?; discuss books, films or television; programmes with you?; listen to classical music with you?; discuss how well you are doing at school?; eat <the main meal> with you around a table?; spend time just talking to you?
School relationships	HBSC	2002	11, 13, and 15	How often have you been bullied at school in the past couple of months?	Likert scale (I haven't been bullied at school the past couple of months, It has only happened once or twice, 2 or 3 times a month, About once a week, Several times a week).
	HBSC	2002	11, 13, and 15	How often have you taken part in bullying another student(s) at school in the past couple of months?	Likert scale (I haven't bullied at school the past couple of months, It has only happened once or twice, 2 or 3 times a month, About once a week, Several times a week).
	PIRLS	2001 and 2006	9 to 10	Did any of these things happen at school during the last month (as far as you know)?	Yes/No: Something was stolen from me ; something was stolen from someone in my class ; I was bullied by another student ; someone in my class was bullied by another student ; I was hit or hurt by another student ; someone in my class was hit or hurt by another student.
	TIMSS	2003 and 2007	10 (2003), 9 and 13 (2007)	Did any of these things happen at school during the last month (as far as you know)?	Yes/No: Something of mine was stolen; I was hit or hurt by another student; I was made to do things by other students; I was made fun of or called names; I was left out of activities.
Antisocial behaviour	ESPAD	2007	15 to 16	Last year, how often? Hit teacher; fight; used weapon; stolen; arson; damaged property; got into trouble with police.	Ordinal scale:
	ESPAD	2007	15 to 16	How often have you? Bullied; hit someone; stolen from someone; damaged someone's property?	Ordinal scale:
	ESPAD	2007	15 to 16	How often has the following been done to you? Bullied; hit; stolen from; damaged your property?	Ordinal scale:
	HBSC	2002	11, 13, and 15	During the past 12 months, how many times were you in a physical fight?	Likert scale (I have not been in a physical fight in the past 12 months, 1 time, 2 times, 3 times, 4 times or more).

Subjective perceptions and opinions

Sub-domain	Source	Most recent date	Age or group covered	Subject or question	Item detail
Aspirations	CIVED	1999	14	How many years of further education do you expect to complete after this year?	Ordinal list in two year brackets.
	PISA	2000	15	What kind of job do you expect to have when you are about 30 years old?	Open ended:
	PISA	2003	15	Grade that the pupil expects to complete	Ordinal list:
	TIMSS	2007	13	How far ISCED level do you expect to go?	
Body image	HBSC	2002	11, 13, and 15	Do you think you are….?	Likert scale (Very good looking, quite, about average, not very, or not at all good looking).
	HBSC	2002	11, 13, and 15	Do you think your body is……..?	Likert scale (Much too thin; A bit too thin; About the right size; A bit too fat; Much too fat).
Education	PIRLS	2001 and 2006	9 to 10	What do you think about reading? Tell how much you agree with each of these statements.	Likert scale (Agree a lot, Agree a little, Disagree a little, Disagree a lot): I read only if I have to ; I like talking about books with other people ; I would be happy if someone gave me a book as a present ; I think reading is boring; I need to read well for my future ; I enjoy reading.
	PISA	2006	15	How important do you think it is to do well in science/maths/language?	Yes/No:
	PISA	2006	15	How much do you agree with the statement below about learning science?	Likert scale (???): I generally have fun when I am learning; I like reading about science; I am happy doing science problems; I enjoy acquiring new knowledge; I am interested in learning about science.
	PISA	2006	15	How much interest about learning? (physics, chemistry, biology, astronomy, geology, experiments)	Likert scale: High interest; medium interest; low interest; no interest
	PISA	2006	15	Subjects at my school provide basic skills for science-related career, different career	Likert scale: Strongly agree; Agree; Disagree; Strongly disagree
	PISA	2000	15	How much do you disagree or agree with these statements about reading? I cannot sit still and read for more than a few minutes; I read only to get information that I need; I enjoy going to a bookstore or a library; For me, reading is a waste of time; I feel happy if I receive a book as a present; I find it hard to finish books; I like talking about books with other people; Reading is one of my favourite hobbies; I read only if I have to.	Likert scale: Strongly agree; Agree; Disagree; Strongly disagree
	TIMSS	2003	10	How much do you agree with these statements about learning mathematics?	Likert scale (Agree a lot, Agree a little, Disagree a little, Disagree a lot): Same questions asked for science. I usually do well in mathematics; I would like to do more mathematics in school; Mathematics is harder for me than for many of my classmates; I enjoy learning mathematics; I am just not good at mathematics; I learn things quickly in mathematics; I practice adding, subtracting, multiplying, and dividing without using a calculator; I work on fractions and decimals; I measure things in the classroom and around the school; I make tables, charts, or graphs; I learn about shapes such as circles, triangles, and rectangles; I work with other students in small groups; I explain my answers; I listen to the teacher talk; I work problems on my own; I use a calculator.
	TIMSS	2007	9 to 13	How much do you agree - maths? Do well, want to do more, harder for me, can learn things quickly, is boring, like.	Likert scale: asked for mathematics and science.
	TIMSS	2007	9 to 13	Like being in school; Students in school try to do their best; Teachers want student to do their best	Likert scale: Strongly agree; Agree; Disagree; Strongly disagree
School life	HBSC	2002	11, 13, and 15	Here are some statements about the students in your class(es). Please show how much you agree or disagree with each one.	Likert scale (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree): The students in my class(es) enjoy being together; Most of the students in my class(es) are kind and helpful; Other students accept me as I am.
	HBSC	2002	11, 13, and 15	How do you feel about school at present?	Likert scale (I like it a lot, I like it a bit, I don't like it very much, I don't like it at all).
	HBSC	2002	11, 13, and 15	How pressured do you feel by the schoolwork you have to do?	Likert scale (Not at all, A little, Some, A lot).
	HBSC	2002	11, 13, and	In your opinion, what does your class teacher(s) think about	Likert scale (Very good, good average, below

			15	your school performance compared to your classmates?	average).
	PIRLS	2001, 2006	9 to 10	What do you think about your school? Tell how much you agree with these statements.	Likert scale (Agree a lot, Agree a little, Disagree a little, Disagree a lot): I feel safe when I am at school; I like being in school; I think that students in my school work hard; I think that teachers in my school care about me; I think that teachers in my school want students to work hard.
	PISA	2000	15	My school is a place where:	Likert scale (???):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; I feel lonely; I do not want to go; I often feel bored.
	PISA	2003	15	Thinking about what you have learned in school: To what extent do you agree with the following statements?	Likert scale (???): School has done little to prepare me for adult life when I leave school; School has been a waste of time.; School has helped give me confidence to make decisions; School has taught me things which could be useful in a job.
	PISA	2003	15	Which of the following are reasons why you attend this school?	Tick box: This is the local school for students who live in this area; This school is known to be a better school than others in the area; This school offers specific study programmes; This school has a particular religious philosophy; Previously, family members attended this school; Other reasons.
	TIMSS	2003	10	How much do you agree with these statements about your school?	Likert scale (Agree a lot, Agree a little, Disagree a little, Disagree a lot): I like being in school; I think that most students in my school try to do their best; I think that most teachers in my school care about the students; I think that most teachers in my school want students to do their best.
Health	HBSC	2002	11, 13, and 15	Would you say your health is……?	Likert scale (Excellent, Good, Fair, poor).
	ESS	2008	15+	How is your health in general?	Ordinal scale: Very good; good; fair; bad; very bad; (Don't know).
Life satisfaction	ESPAD	2007	15 to 16	How satisfied with? Finance; health; yourself	Likert scale:
	ESS	2008	15+	How happy are you with your life?	Ordinal scale:
	HBSC	2002	11, 13, and 15	Here is a picture of a ladder. The top of the ladder '10' is the best possible life for you and the bottom '0' is the worst possible life for you. In general, where on the ladder do you feel you stand at the moment? Tick the box next to the number that best describes where you stand.	Discrete scale: Cantrill Ladder 10 says 'Best possible life', 0 says worst possible life'.
Personal relationships	EQLS	2007	15+	How often have the following happened? (too tired for household jobs/ difficult to fulfil family responsibility/ difficult to concentrate on work due to family responsibilities)	Ordinal scale:
	ESPAD	2007	15 to 16	Happy with parents?	Likert scale:
	ESS	2008	15+	Trust in others?	Likert scale:
	ESS	2008	15+	Would people like to take advantage of you?	Likert scale:
	ESS	2008	15+	People try to be helpful most of the time/	Likert scale:
	PISA	2000	15	How much do you disagree or agree with each of the following statements about teachers at your school?	Likert scale: Most of my teachers treat me fairly; If I need extra help, I will receive it from my teachers; Most of my teachers really listen to what I have to say; Most teachers are interested in students' well-being; Students get along well with most teachers.
Political opinion	ICCS	2009	13 to 14	Do you agree that men and women should have equal opportunity?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
	ICCS	2009	13 to 14	Do you agree that men and women should have the same rights in every way?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
	ICCS	2009	13 to 14	Do you agree that women should stay out of politics?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
	ICCS	2009	13 to 14	Do you agree that when there are not enough jobs men should have more right to jobs than women?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
	ICCS	2009	13 to 14	Do you agree that men and women should get equal pay in the same jobs?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
	ICCS	2009	13 to 14	Do you think men are better qualified to be political leaders?[2]	Likert scale: Strongly disagree; Disagree; Agree; Strongly Agree
Satisfaction with locality	HBSC	2002	11, 13, and 15	Do you think that the area in which you live is a good place to live?	Likert scale (yes it's really good, yes it's good, ok, it's not very good, no, it's not good at all).
	HBSC	2002	11, 13, and 15	Please say how you feel about these statements about the area in which you live.	Likert scale (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree): people say 'hello' and often stop to

					talk to each other in the street; it is safe for younger children to play outside during the day; you can trust people around here; there are good places to spend your free time (leisure centres, parks and shops); I can ask for help or a favour from neighbours; most people around here would try to take advantage of you if they could.
Subjective material situation	EQLS	2007	15+	How easily is your household able to make ends meet?	Likert scale:
	ESPAD	2007	15 to 16	How well off is your family compared to other families in neighbourhood?	Ordinal scale:
	HBSC	2002	11, 13, and 15	How well off do you think your family is?	Ordinal List (Very well off, Quite well off, Average, Not very well off, Not at all well off).

1. Questions that allow for information on 0-17 year olds even though not directly asked to children.

2. Questions derived based on study rather than actual questionnaire.

ANNEX 6: MISSING DATA FROM PISA 2009 BY SEX AND MIGRANT STATUS

Below are two tables representing the missing values in PISA 2009 across eight selected indicators. Table 3 looks at missing values by sex, and Table 4, missing values by migrant status. Boys tend to report missing values more often than girls, and migrant pupils report more missing values than non-migrant pupils.

Items with information about parents (education and living at home), as well as whether grades have been repeated, have the most missing values. Questions about the home environment, attendance of pre-school and classroom environment are rarely missed by respondents.

Missing data on selected PISA 2009 indicators by sex

Country		Proportion of missing data							
		Father's education level	Mother's education level	ISCED 2 repeated	How many books at home?	Father lives at home	Mother lives at home	Went to preschool	Cannot work well in lessons
Australia	Female	8.3	3.4	6.1	0.6	6.8	2.2	0.8	0.9
	Male	8.8	5.6	7.5	1.2	5.9	2.7	1.0	1.6
Austria	Female	8.1	4.5	2.1	0.7	2.9	1.1	0.2	1.3
	Male	9.4	7.1	2.4	1.3	3.5	2.2	0.4	2.6
Belgium	Female	10.2	6.9	9.0	0.6	4.3	1.3	1.1	1.3
	Male	9.9	7.7	11.0	1.3	3.9	1.8	1.1	1.4
Bulgaria	Female	5.0	3.0	3.3	1.5	6.9	3.4	0.4	2.7
	Male	7.3	4.5	4.9	2.2	9.0	5.7	1.1	4.6
Canada	Female	5.1	2.0	2.7	0.6	4.1	1.5	1.4	0.9
	Male	5.2	2.9	3.5	1.5	3.7	1.9	1.9	1.3
Chile	Female	7.5	3.3	11.1	0.3	5.1	2.4	1.0	0.4
	Male	6.6	3.7	14.9	0.9	5.4	3.3	1.7	0.6
Czech Republic	Female	5.4	2.2	1.1	0.3	1.8	0.3	0.1	0.2
	Male	6.3	3.7	1.3	0.9	1.8	0.9	0.4	0.6
Denmark	Female	8.3	5.6	4.4	1.6	4.3	1.8	1.4	1.2
	Male	8.2	6.8	5.7	2.0	4.3	2.1	1.1	1.8
Estonia	Female	9.1	1.1	4.4	0.3	8.9	2.5	0.1	0.1
	Male	7.6	1.5	5.7	0.6	8.7	3.6	0.2	0.4
Finland	Female	4.0	1.4	0.3	0.3	3.0	1.3	0.4	0.2
	Male	4.9	2.9	0.6	1.1	4.6	3.1	0.9	1.0
France	Female	11.2	5.7	6.2	0.5	4.0	1.1	1.1	0.8
	Male	11.5	8.7	9.6	1.4	4.4	1.5	1.9	2.3
Germany	Female	13.6	8.5	3.5	0.9	3.9	1.0	1.4	1.5
	Male	12.5	10.9	3.5	1.6	4.0	1.4	0.8	2.0
Greece	Female	2.2	1.6	1.7	0.3	4.2	1.9	0.5	0.2
	Male	1.9	1.7	1.6	0.6	2.7	1.5	0.9	0.9
Hungary	Female	5.8	1.6	1.9	0.2	2.4	0.9	0.3	0.3
	Male	5.5	2.0	1.8	0.2	2.7	1.1	0.7	0.3
Iceland	Female	2.3	0.7	0.7	0.4	3.8	1.2	0.1	0.8
	Male	2.8	2.3	1.4	1.0	3.5	0.9	0.3	0.9
Ireland	Female	5.7	2.6	7.1	0.4	5.2	1.8	0.6	0.4
	Male	5.9	4.3	7.9	1.1	5.5	2.5	0.9	1.1
Israel	Female	8.8	4.5	2.4	1.1	3.1	1.8	2.3	2.0
	Male	7.7	5.9	3.2	2.5	3.7	2.1	2.6	3.2
Italy	Female	3.6	2.3	6.3	0.4	3.6	1.0	0.5	0.5
	Male	3.6	2.9	8.7	0.8	3.6	1.7	1.0	0.8
Japan	Female	9.4	3.5	0.0	0.3	2.1	0.4	0.1	0.7
	Male	11.3	6.4	0.0	0.9	2.3	0.7	0.4	0.5
Korea	Female	3.3	2.8	0.0	0.1	4.3	2.3	0.4	0.3
	Male	3.7	3.3	0.0	0.2	3.1	2.2	0.3	0.6
Latvia	Female	11.6	3.0	3.4	0.2	8.0	2.1	0.2	0.2
	Male	11.0	2.7	5.7	0.6	8.9	3.1	0.7	0.7
Lithuania	Female	7.1	1.9	2.2	0.2	5.4	1.2	0.3	0.2
	Male	8.4	3.3	5.1	0.9	7.6	1.7	0.8	0.7
Luxembourg	Female	11.8	8.2	8.2	0.5	3.3	0.9	0.6	3.0
	Male	12.3	11.2	9.1	0.7	4.1	1.4	0.6	4.2
Mexico	Female	5.6	1.7	8.2	2.0	10.7	4.1	0.6	1.2
	Male	4.6	1.7	10.8	2.2	10.8	5.2	0.8	1.4
Netherlands	Female	5.1	2.2	7.6	0.7	3.2	0.9	0.5	0.4
	Male	6.3	4.4	7.3	1.5	3.0	1.1	0.9	0.8
New Zealand	Female	15.5	11.1	2.8	0.2	1.0	0.6	0.3	0.4
	Male	14.8	12.7	3.0	0.6	1.0	0.7	0.7	1.0
Norway	Female	6.0	2.4	0.0	0.6	4.4	1.0	0.1	0.4
	Male	5.9	4.0	0.0	0.8	3.9	1.9	0.2	0.7
Poland	Female	5.3	2.3	1.3	0.2	0.2	0.2	0.1	0.2
	Male	5.3	3.7	2.2	0.5	0.6	0.6	0.1	0.7
Portugal	Female	3.6	1.2	10.3	0.2	6.5	2.0	0.4	0.3
	Male	3.3	1.7	15.1	0.6	5.6	1.9	0.8	0.8
Romania	Female	1.7	0.9	3.0	0.4	4.2	2.6	0.1	0.7
	Male	1.8	1.6	4.5	0.7	4.7	2.5	0.3	1.3
Slovak Republic	Female	3.5	0.9	4.0	0.1	3.1	0.9	0.1	0.5
	Male	2.9	1.5	6.3	0.7	4.5	1.8	0.5	0.8
Slovenia	Female	5.5	1.9	0.1	0.5	3.5	1.4	0.2	0.6
	Male	5.6	3.3	0.6	1.4	3.8	2.4	0.6	2.0
Spain	Female	5.6	2.5	2.5	0.7	4.1	1.2	1.3	0.8
	Male	5.4	3.7	2.9	1.0	3.7	1.5	2.0	1.4
Sweden	Female	8.3	4.5	3.3	0.3	5.6	2.5	0.6	0.5
	Male	8.5	6.8	3.7	1.3	6.2	3.4	1.0	1.5
Switzerland	Female	5.8	4.4	8.9	0.3	3.2	1.1	0.7	0.7
	Male	7.4	6.2	10.8	0.9	4.4	1.9	1.3	1.2
Turkey	Female	3.7	3.4	0.0	0.7	4.5	2.5	0.6	0.8
	Male	3.9	4.3	0.0	1.4	5.5	3.6	1.6	1.1
United Kingdom	Female	12.3	6.0	1.7	0.7	7.6	2.0	0.6	0.6
	Male	14.0	10.1	1.9	1.5	6.7	2.1	1.4	1.2
United States	Female	6.2	1.7	6.8	0.7	8.1	2.9	0.5	2.4
	Male	6.0	2.3	10.1	1.5	8.9	4.2	0.9	2.7
Average	**Female**	**6.9**	**3.4**	**3.9**	**0.5**	**4.5**	**1.6**	**0.6**	**0.8**
	Male	**7.1**	**4.7**	**5.1**	**1.1**	**4.7**	**2.2**	**0.9**	**1.4**

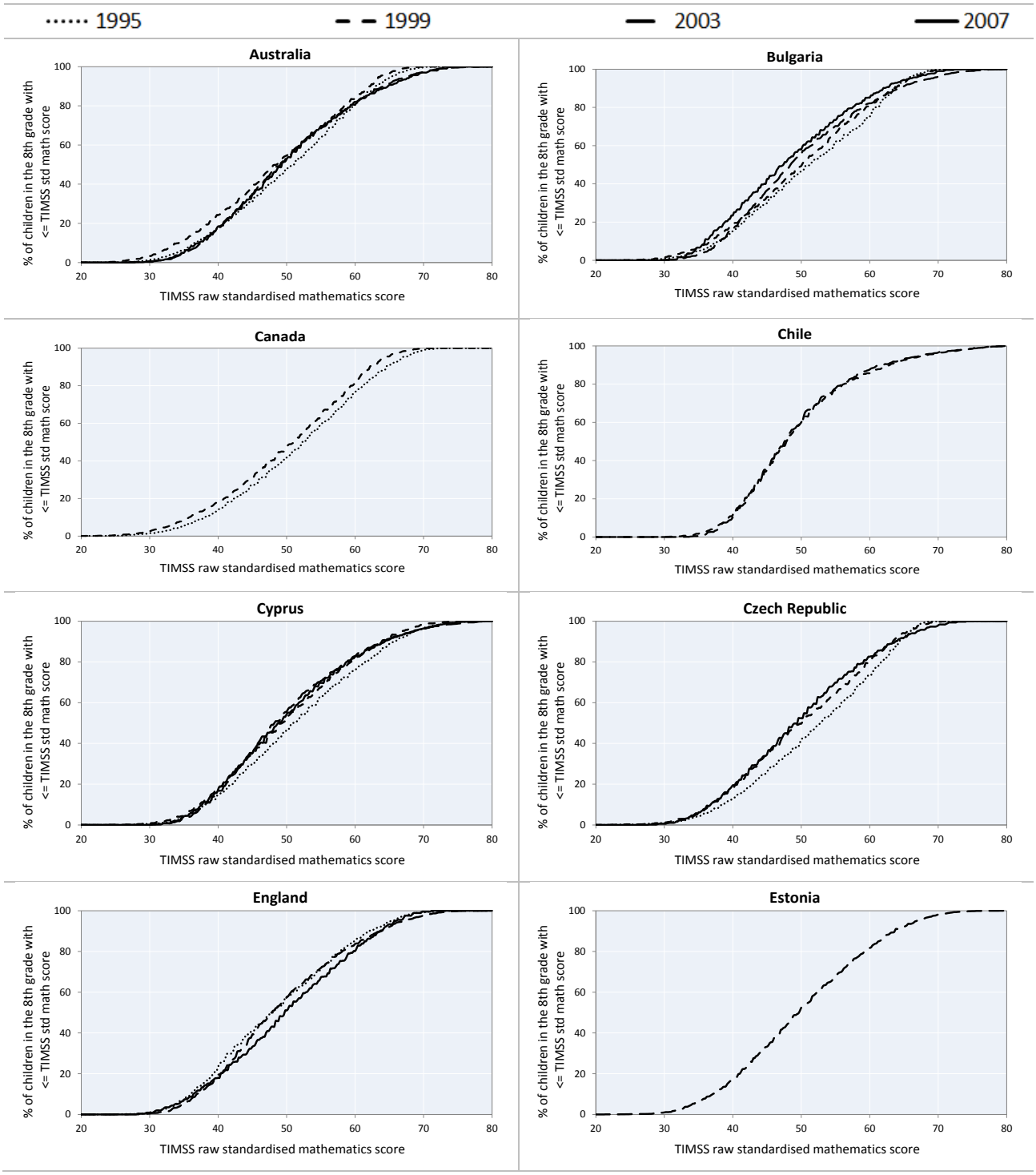
Missing data on selected PISA 2009 indicators by migrant status

Country		Proportion of missing data							
		Father's education level	Mother's education level	ISCED 2 repeated	How many books at home?	Father lives at home	Mother lives at home	Went to preschool	Cannot work well in lessons
Australia	Country of test	8.4	4.2	6.6	0.9	6.2	2.3	0.8	1.1
	Other country	8.7	5.4	7.8	0.6	6.8	2.8	1.3	1.1
Austria	Country of test	8.4	5.3	2.1	0.8	3.0	1.5	0.3	1.8
	Other country	9.4	7.5	2.9	0.3	2.9	1.9	0.5	2.7
Belgium	Country of test	9.2	6.6	9.1	0.7	3.8	1.4	0.9	1.2
	Other country	14.4	11.0	16.7	1.4	5.3	2.8	1.7	2.2
Bulgaria	Country of test	5.5	3.2	3.5	1.4	7.3	4.1	0.4	3.0
	Other country	5.5	3.6	1.8	1.8	7.3	1.8	1.8	7.3
Canada	Country of test	4.8	2.1	2.8	0.8	3.7	1.5	1.4	0.8
	Other country	9.1	6.3	6.9	3.8	7.1	4.1	4.9	4.0
Chile	Country of test	6.9	3.5	13.1	0.6	5.1	2.7	1.4	0.4
	Other country	10.3	4.4	7.4	0.0	10.3	8.8	0.0	0.0
Czech Republic	Country of test	5.8	2.9	1.1	0.6	1.7	0.6	0.3	0.4
	Other country	6.5	2.8	1.9	0.0	2.8	0.9	0.0	1.9
Denmark	Country of test	7.1	5.3	4.6	1.7	4.0	1.8	1.3	1.4
	Other country	17.7	13.2	9.2	2.6	6.8	3.0	1.5	2.4
Estonia	Country of test	8.3	1.3	5.0	0.4	8.7	3.0	0.1	0.3
	Other country	7.1	1.2	7.1	2.4	15.3	4.7	0.0	0.0
Finland	Country of test	4.2	2.0	0.4	0.7	3.7	2.1	0.6	0.6
	Other country	9.6	4.5	1.9	0.6	8.3	3.2	0.0	1.3
France	Country of test	11.0	7.0	7.6	0.9	4.0	1.2	1.4	1.4
	Other country	14.6	8.5	10.8	0.0	5.6	0.9	2.8	1.9
Germany	Country of test	12.4	9.0	3.2	1.2	4.0	1.1	0.9	1.5
	Other country	30.8	25.4	8.3	0.6	5.0	1.2	3.8	4.1
Greece	Country of test	1.8	1.4	1.5	0.4	3.4	1.8	0.7	0.5
	Other country	3.7	4.0	2.9	0.5	3.4	0.5	0.5	1.3
Hungary	Country of test	5.6	1.8	1.8	0.2	2.6	1.0	0.5	0.3
	Other country	3.5	3.5	5.9	0.0	1.2	1.2	1.2	0.0
Iceland	Country of test	2.0	1.2	0.9	0.5	3.5	1.0	0.1	0.8
	Other country	5.4	2.3	1.8	1.8	4.5	0.5	0.0	0.0
Ireland	Country of test	5.3	3.5	7.2	0.7	5.0	2.0	0.6	0.6
	Other country	8.4	3.0	9.0	0.9	7.7	3.2	1.3	1.3
Israel	Country of test	7.0	4.4	2.3	1.6	3.3	2.0	2.1	2.4
	Other country	18.3	11.0	5.4	0.8	3.9	0.4	4.0	2.1
Italy	Country of test	3.3	2.4	7.2	0.5	3.5	1.3	0.7	0.5
	Other country	7.5	5.0	11.1	1.3	5.7	2.1	1.9	1.6
Japan	Country of test	10.3	4.9	0.0	0.6	2.2	0.6	0.2	0.5
	Other country	12.0	16.0	0.0	0.0	0.0	0.0	0.0	4.0
Korea	Country of test	3.5	3.0	0.0	0.1	3.6	2.2	0.3	0.4
	Other country	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Latvia	Country of test	11.3	2.8	4.5	0.4	8.4	2.6	0.5	0.5
	Other country	10.7	2.7	1.3	0.0	5.3	1.3	0.0	0.0
Lithuania	Country of test	7.4	2.4	3.6	0.5	6.2	1.3	0.5	0.5
	Other country	13.3	0.0	6.7	0.0	20.0	3.3	0.0	0.0
Luxembourg	Country of test	11.6	9.5	7.5	0.3	3.6	1.0	0.4	3.4
	Other country	13.1	9.2	13.5	1.5	3.7	1.4	1.5	4.1
Mexico	Country of test	4.9	1.6	8.9	2.1	10.4	4.5	0.6	1.1
	Other country	4.8	3.7	17.2	1.5	14.7	5.9	1.0	3.2
Netherlands	Country of test	4.9	2.8	7.2	1.0	2.9	1.0	0.6	0.5
	Other country	20.9	13.0	10.7	0.9	7.0	2.3	0.9	1.9
New Zealand	Country of test	16.5	12.6	3.0	0.4	0.9	0.5	0.4	0.6
	Other country	10.1	9.0	2.3	0.3	1.1	0.9	0.5	0.7
Norway	Country of test	5.4	2.9	0.0	0.7	3.9	1.3	0.1	0.5
	Other country	12.9	6.4	0.0	0.0	7.2	2.8	0.0	0.4
Poland	Country of test	5.2	3.0	1.7	0.3	0.4	0.4	0.1	0.4
	Other country	5.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0
Portugal	Country of test	3.3	1.3	12.2	0.4	5.7	1.8	0.6	0.5
	Other country	5.5	2.5	15.5	0.6	10.5	4.0	1.1	0.0
Romania	Country of test	1.6	1.2	3.6	0.4	4.4	2.5	0.2	0.9
	Other country	10.3	0.0	6.9	0.0	6.9	6.9	0.0	0.0
Slovak Republic	Country of test	3.0	1.0	5.0	0.4	3.6	1.3	0.3	0.5
	Other country	10.8	5.4	5.4	0.0	8.1	0.0	0.0	2.7
Slovenia	Country of test	5.1	2.4	0.2	0.8	3.5	1.7	0.3	1.0
	Other country	7.6	2.8	0.0	2.8	4.8	4.1	0.7	5.5
Spain	Country of test	5.0	3.0	2.3	0.7	3.6	1.2	1.3	0.9
	Other country	8.4	2.9	6.1	0.8	6.4	2.7	4.2	1.9
Sweden	Country of test	7.7	5.3	3.0	0.7	5.6	2.8	0.7	0.9
	Other country	16.4	8.6	10.5	0.8	6.6	3.1	1.2	1.6
Switzerland	Country of test	5.9	4.8	8.8	0.6	3.7	1.3	0.8	0.8
	Other country	9.8	7.4	16.5	0.4	4.2	2.4	1.7	1.1
Turkey	Country of test	3.6	3.7	0.0	1.0	4.8	2.9	1.1	0.9
	Other country	17.1	8.6	0.0	0.0	8.6	5.7	2.9	0.0
United Kingdom	Country of test	12.9	7.9	1.6	1.0	7.0	2.0	0.9	0.8
	Other country	14.4	8.7	3.5	1.0	8.4	2.0	1.3	1.7
United States	Country of test	5.8	1.7	8.3	0.9	8.0	3.3	0.4	2.1
	Other country	7.9	2.4	8.4	0.8	11.0	3.1	0.8	3.9
Average	**Country of test**	**6.6**	**3.8**	**4.3**	**0.7**	**4.4**	**1.8**	**0.7**	**1.0**
	Other country	**10.6**	**6.2**	**6.4**	**0.8**	**6.4**	**2.5**	**1.2**	**1.8**

ANNEX 7: COMPARISONS OF DISTRIBUTION IN MATHEMATICS (TIMSS) AND READING (PIRLS) ACROSS SURVEY WAVES

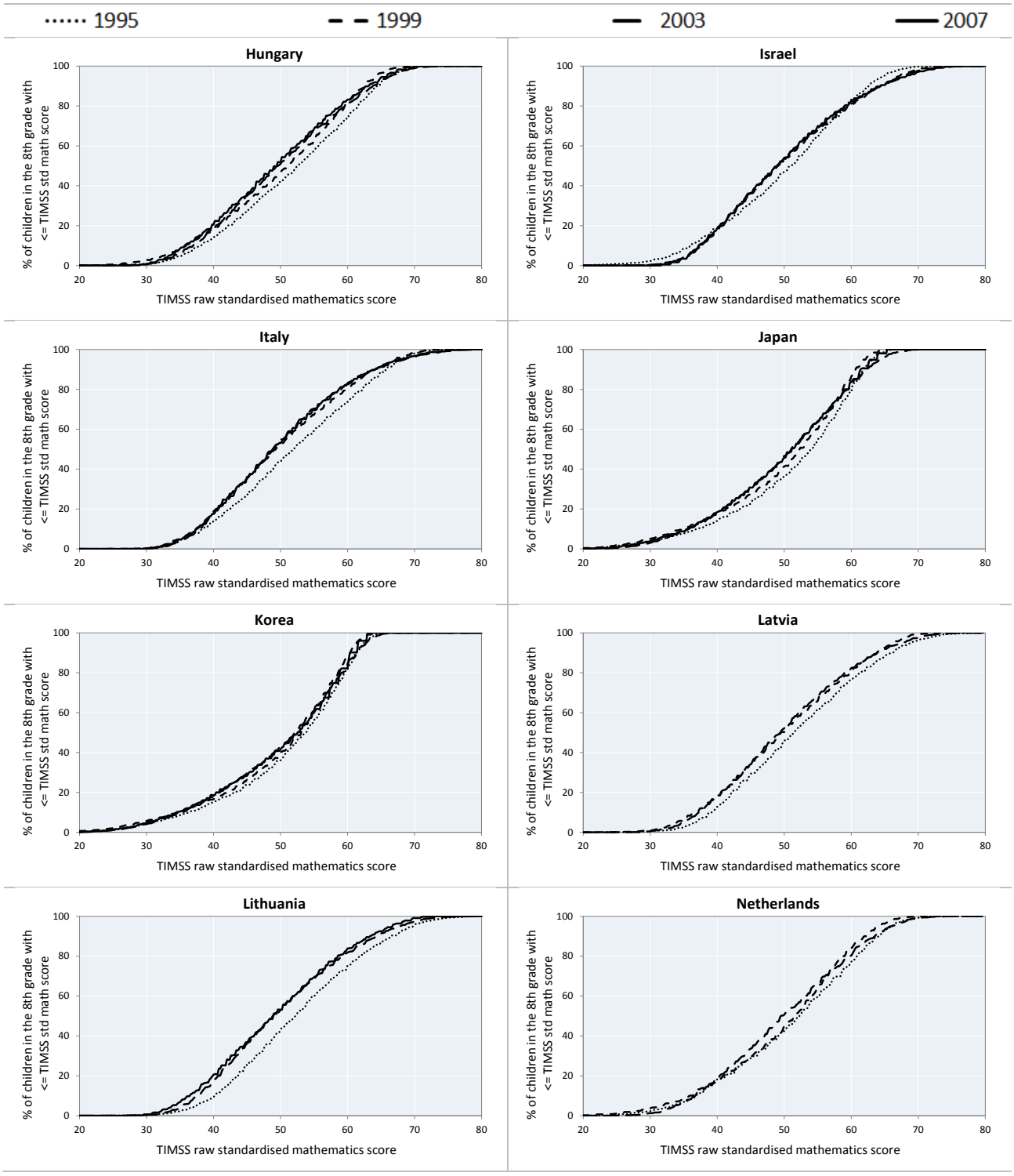
Annex Figure 1: The standardised mathematics scores are normally distributed in most OECD and EU countries

The distribution of raw standardised mathematics scores among 8th grade pupils in TIMSS 1995, 199, 2003 and 2007



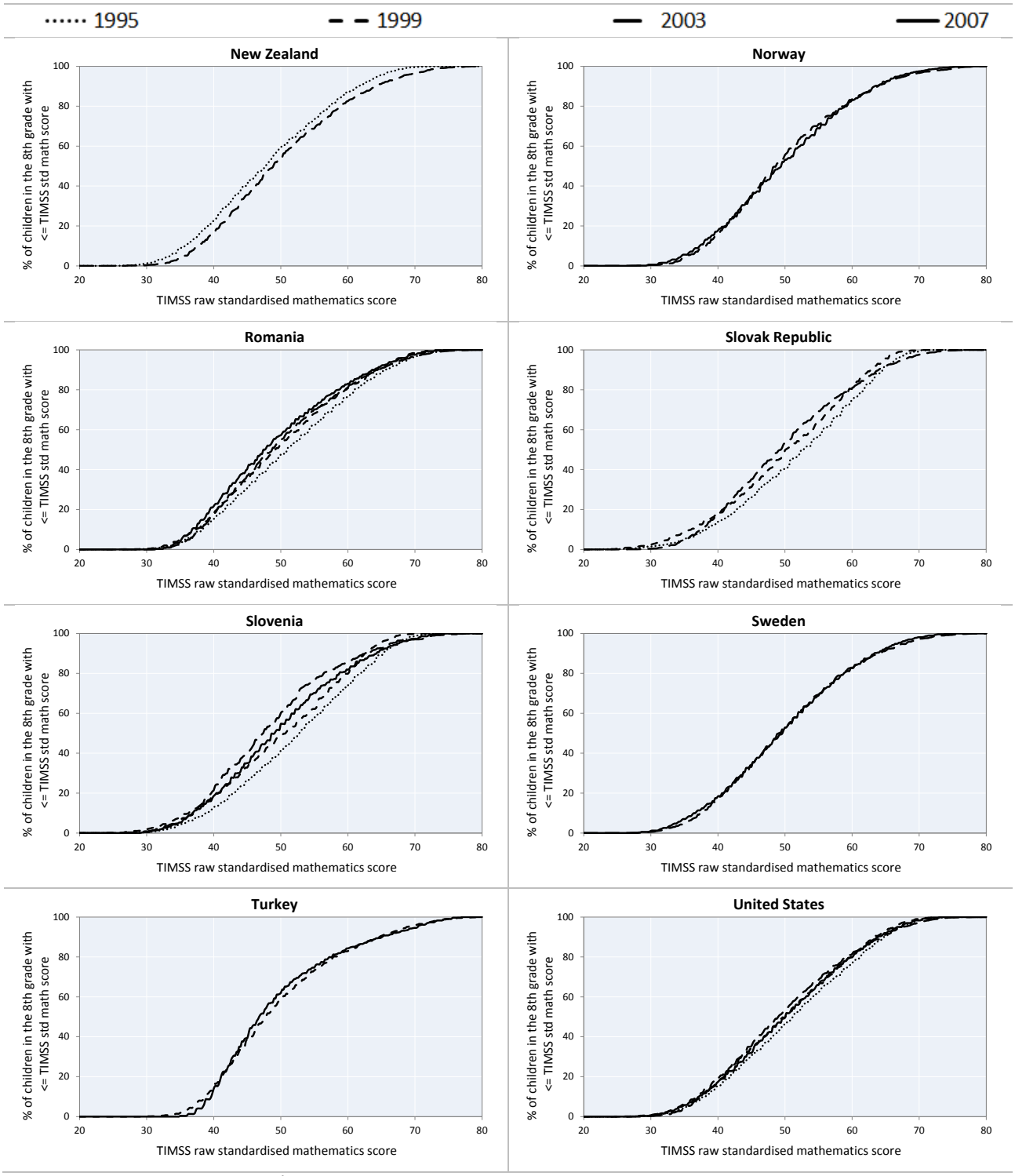
Annex Figure 1: The standardised mathematics scores are normally distributed in most OECD and EU countries (cont.)

The distribution of raw standardised mathematics scores among 8th grade pupils in TIMSS 1995, 199, 2003 and 2007



Annex Figure 1: The standardised mathematics scores are normally distributed in most OECD and EU countries (cont.)

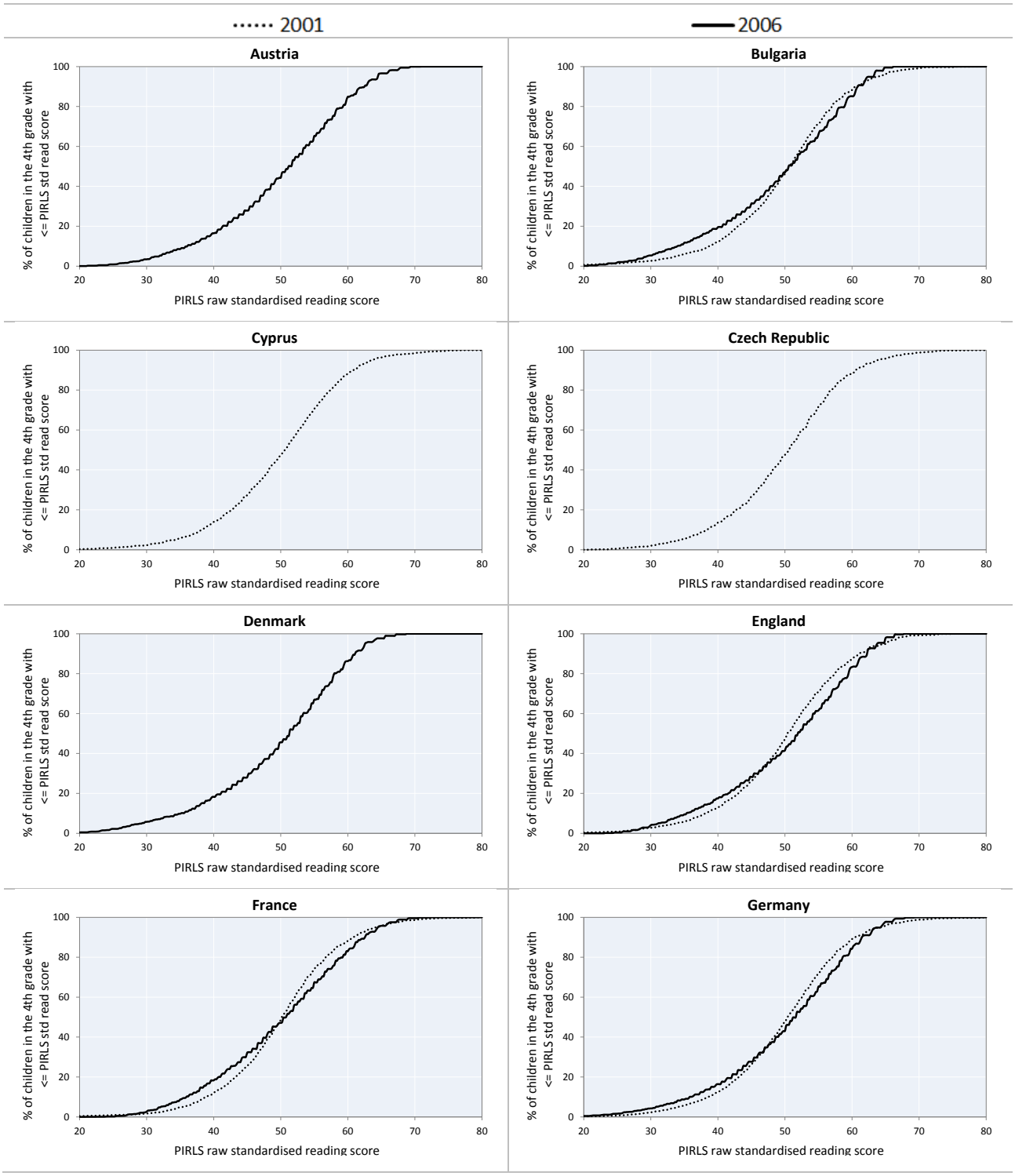
The distribution of raw standardised mathematics scores among 8th grade pupils in TIMSS 1995, 199, 2003 and 2007



Source: TIMSS 1995, 1999, 2003 and 2007.

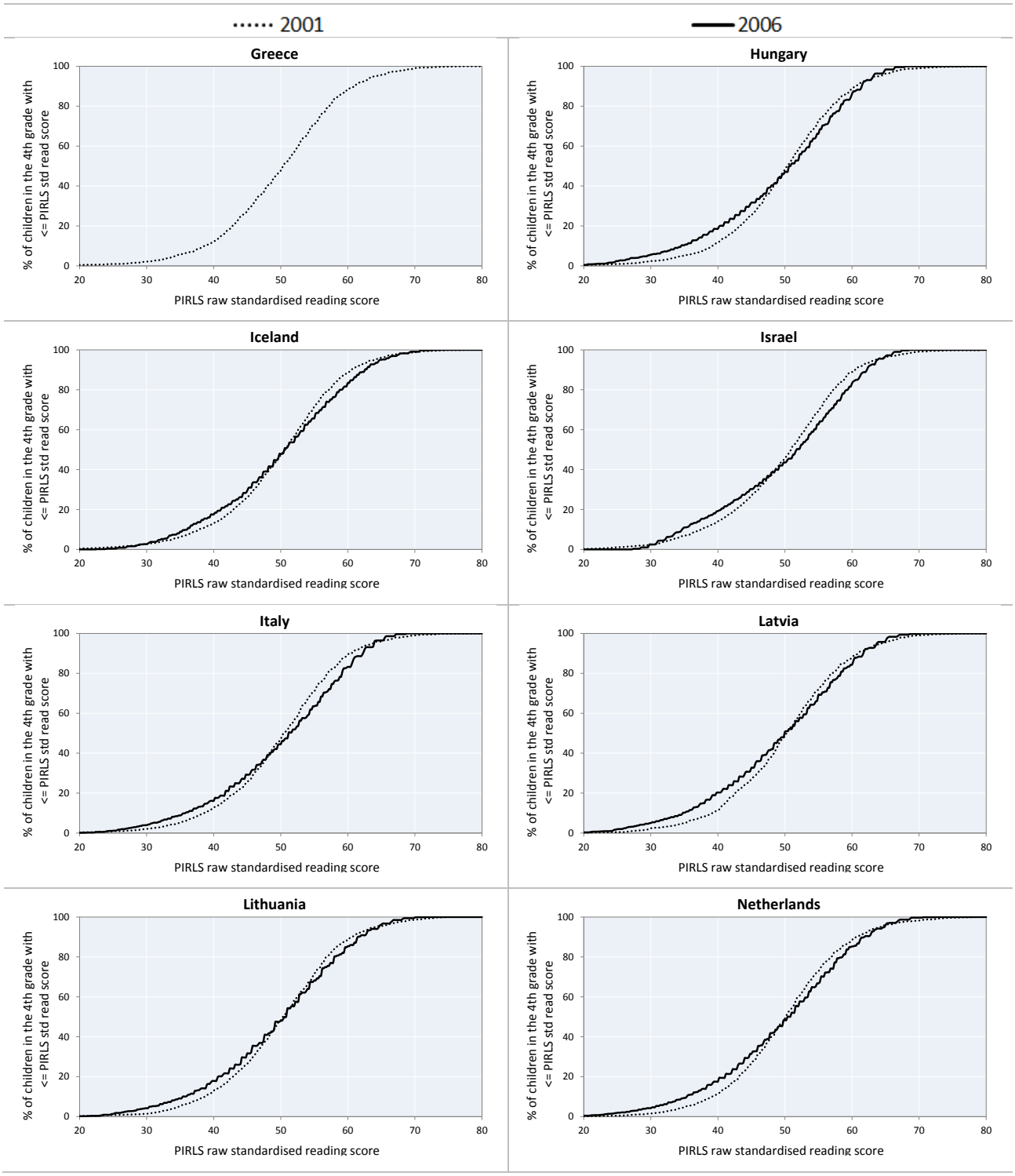
Annex Figure 2: The standardised reading scores are normally distributed in most OECD and EU countries

The distribution of raw standardised reading scores among 4th grade pupils in PIRLS 2001 and 2006



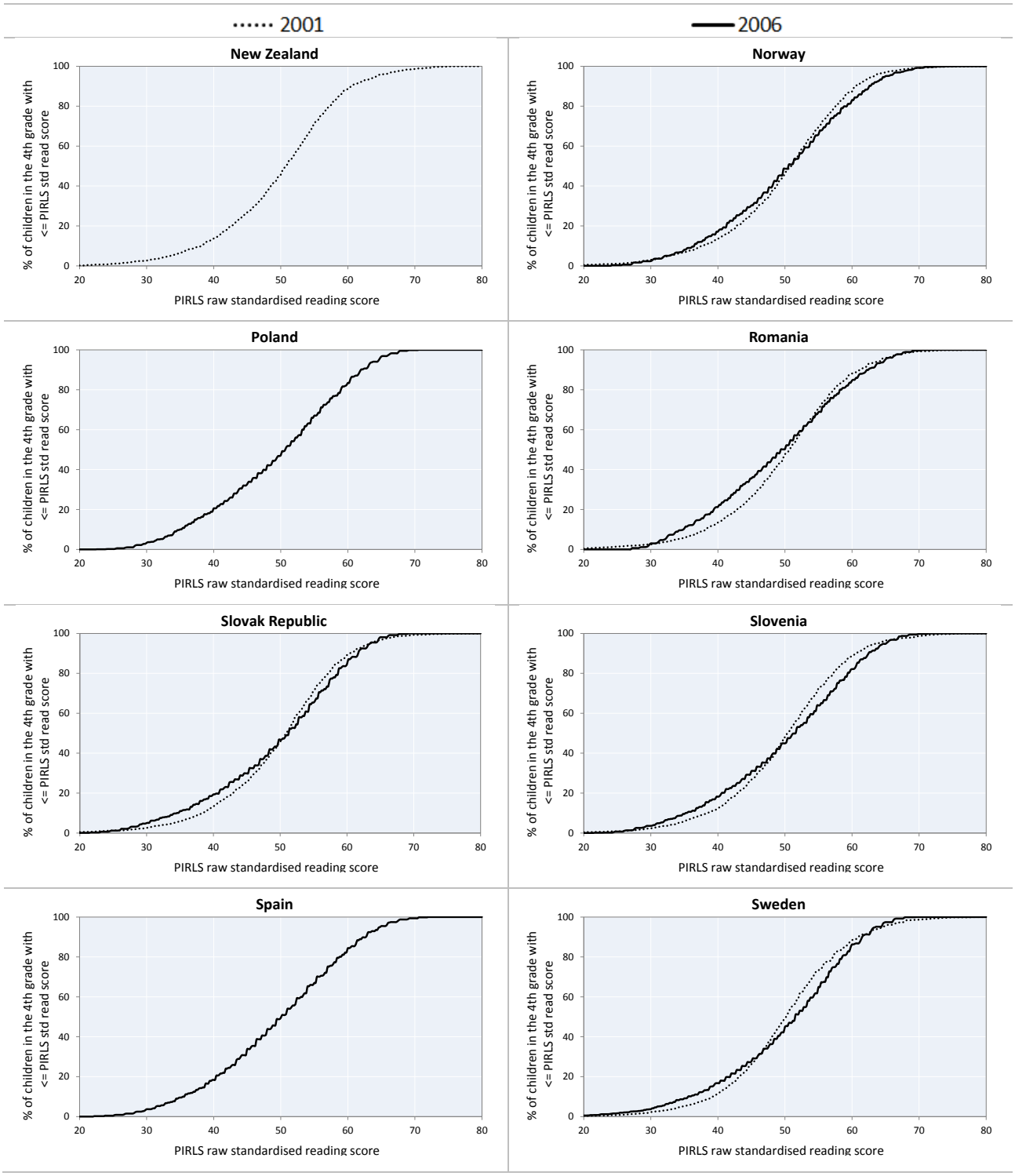
Annex Figure 2: The standardised reading scores are normally distributed in most OECD and EU countries (cont.)

The distribution of raw standardised reading scores among 4th grade pupils in PIRLS 2001 and 2006



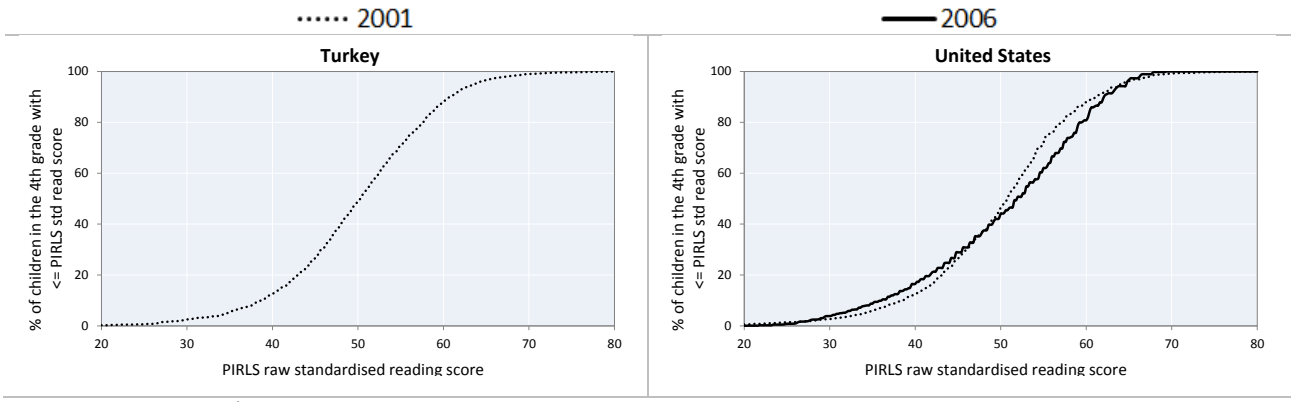
Annex Figure 2: The standardised reading scores are normally distributed in most OECD and EU countries (cont.)

The distribution of raw standardised reading scores among 4th grade pupils in PIRLS 2001 and 2006



Annex Figure 2: The standardised reading scores are normally distributed in most OECD and EU countries (cont.)

The distribution of raw standardised reading scores among 4th grade pupils in PIRLS 2001 and 2006



Source: PIRLS 2001 and 2006.

ANNEX 8: RATES OF CHILDREN IN SPECIAL NEEDS EDUCATION

Table A8.1 Special Educational Needs Breakdown by settings varies a lot across countries

Children with SEN in different settings as a percentage of total number of children with SEN, various years

Country	2008 round			2010 round		
	Special schools	Special classes in mainstream schools	Inclusive settings	Special schools	Special classes in mainstream schools	Inclusive settings
Austria	36.9	7.7	55.4	41.3	3.4	55.3
Czech Republic	40.9	11.7	47.3	41.9	9.8	48.4
Denmark	21.8	70.2	8.0	37.8	56.3	5.9
Estonia	18.1	7.1	74.7	32.2	14.0	53.8
Finland	16.1	32.7	51.2	14.9	32.0	53.1
France	22.4	46.7	30.9	21.2	45.2	33.6
Germany	84.9		15.1	83.2		16.7
Greece	26.2	73.8		25.0	75.0	
Hungary	50.0		50.0	46.7		53.3
Iceland	7.5	17.9	74.6	1.3	3.3	95.4
Ireland	39.5	60.5		14.7	7.0 (2)	78.3
Italy	0.4	n.a.	99.6	m.	m.	m.
Luxembourg	52.8		47.2	48.3		51.7
Netherlands	67.0	m.	33.0	62.1	m.	37.9
Norway	6.1	m.	93.9	4.0	10.9	85.1
Poland	54.3	n.a.	45.7	46.8	n.a.	53.2
Portugal	8.0	3.0	89.0	7.3	5.9	86.8
Slovenia	14.4	3.3	67.4	26.9	3.8	69.3
Spain	23.7		76.3	16.7		83.3
Sweden	3.6	96.4		3.7	96.3	
Switzerland	m.	m.	m.	39.0	61.0	m.
UK - England	41.4	7.4	51.2	42.6	7.2	50.3
UK - Northern Ireland	m.	m.	m.	29.1	13.4	57.5
UK - Scotland	18.7	5.7	75.6	14.7	3.3	82.1
UK - Wales	21.7	21.1	57.1	23.8	22.0	54.1

Note: m: missing, n.a.: not applicable. The 2008 round refers to academic school year 2006/2007 for Austria, Denmark, Estonia, France, Germany, Hungary, Iceland, Ireland, Luxembourg, Netherlands, Portugal, Spain, Switzerland, UK (Scotland and Wales); 2007/2008 for Czech Republic, Finland, Greece, Norway, Poland, Slovenia, Sweden and UK (England).2010 round refers to academic school year 2007/2008 for Germany, Italy, Portugal, Spain; 2008/2009 for Austria, Denmark, Finland, France, Hungary, Iceland, Ireland, Luxembourg, Sweden, Switzerland and UK (England and Wales); 2009/2010 for Czech Republic, Estonia, Greece, Netherlands, Norway, Poland, Slovenia and UK (Northern Ireland and Scotland).

Source: European Agency for Development in Special Needs Education.

ANNEX 9: FREQUENCY OF RESPONSE BIAS ON KEY INDICATORS: HBSC SURVEYS 2001/02, 2005/06 AND 2009/10

Table A9.1: Frequency and direction of response bias in "Life Satisfaction" query: 2001, 2005 and 2009 HBSC surveys

Country	Sex	Sex Sign	Age	Age Sign	Wealth	Wealth sign	Health	Health sign
Austria	1	+	…	…	2	+	…	…
Belgium	…	…	1	+	3	+	2	-
Czech Rep.	1	+	1	+	2	+	…	…
Canada	…	…	2	+	3	mixed	1	-
Denmark	…	…	1	+	…	…	1	+
Estonia	…	…	…	…	…	…	…	…
Finland	…	…	1	+	…	…	…	…
France	…	…	…	…	2	+	1	-
Germany	…	…	1	+	…	…	…	…
Greece	…	…	1	+	…	…	…	…
Hungary	…	…	1	+	…	…	…	…
Ireland	…	…	…	…	…	…	…	…
Israel	1	+	…	…	2	+	…	…
Latvia	2	+	1	+	…	…	…	…
Lithuania	2	+	2	mixed	2	-	…	…
Luxembourg	2	-	1	+	…	…	2	-
Netherlands	1	+	…	…	1	-	…	…
Norway	…	…	1	+	…	…	2	-
Poland	1	+	2	+	…	…	1	-
Portugal	1	+	3	+	1	+	…	…
Slovenia	…	…	1	+	1	+	…	…
Spain	1	+	1	+	1	+	…	…
Sweden	1	+	…	…	…	…	…	…
Switzerland	…	…	1	+	…	…	…	…
Turkey	…	…	2	+	1	-	…	…
United Kingdom	…	…	3	+	1	+	…	…
United States	…	…	2	+	1	+	…	…

Source: Author's calculations of HBSC survey data for survey waves 2001/02, 2005/06 and 2009/10.

Note: "…" no evidence of response bias. For sex sign, positive means higher male non-response, negative is higher female non-response. For age, positive signs mean older students are significantly more likely to respond. For health and wealth signs, positive means richer or healthier students are significantly more likely to not respond, negative signs mean that and poorer or unhealthier are significantly more likely to be non-respondents. The table presents the results of a binary logit model. Bars indicate the frequency of statistically significant relationships between the identified correlates and likelihood of response to the "Life Satisfaction" question across three waves of the HBSC survey. Statistical significance is defined as $p \leq 0.05$. Greece, Israel, Luxembourg, and Turkey each participated in fewer than three surveys.

Table A9.2: Frequency and direction of response bias in "Weight" query: 2001, 2005 and 2009 HBSC surveys

Country	Sex	Sex Sign	Age	Age Sign	Wealth	Wealth sign	Health	Health sign
Austria	1	-	1	+	…	…	2	-
Belgium	1	-	3	+	3	+	3	-
Czech Rep.	1	-	…	…	2	-	…	…
Canada	3	-	3	+	…	…	3	-
Denmark	1	-	3	+	…	…	3	-
Estonia	1	+	3	+	…	…	2	-
Finland	2	-	…	…	…	…	2	-
France	…	…	2	+	…	…	2	-
Germany	3	-	3	+	1	+	1	+
Greece	2	-	1	+	1	-	…	…
Hungary	…	…	2	+	…	…	2	-
Ireland	3	-	3	+	…	…	2	-
Israel	2	-	2	+	…	…	1	-
Latvia	…	…	3	+	1	+	…	…
Lithuania	3	+	3	+	3	-	…	…
Luxembourg	2	-	1	+	…	…	2	-
Netherlands	…	…	1	+	1	-	1	-
Norway	3	-	3	+	…	…	1	-
Poland	1	-	2	+	…	…	…	…
Portugal	1	-	2	+	…	…	…	…
Slovenia	2	-	…	…	…	…	1	-
Spain	…	…	2	+	…	…	3	-
Sweden	2	-	3	+	…	…	2	-
Switzerland	2	-	1	+	…	…	1	-
Turkey	2	-	1	+	…	…	…	…
United Kingdom	3	-	3	+	…	…	3	-
United States	1	-	3	+	1	+	1	+

Source: Author's calculations of HBSC survey data for survey waves 2001/02, 2005/06 and 2009/10.

Note: "…" no evidence of response bias. For sex sign, positive means higher male non-response, negative is higher female non-response. For age, positive signs mean older students are significantly more likely to respond. For health and wealth signs, positive means richer or healthier students are significantly more likely to not respond, negative signs mean that and poorer or unhealthier are significantly more likely to be non-respondents. For health The table presents the results of a binary logit model. Bars indicate the frequency of statistically significant relationships between the identified correlates and likelihood of response to the "weight" question across three waves of the HBSC survey. Statistical significance is defined as $p \leq 0.05$. Greece, Israel, Luxembourg, and Turkey each participated in fewer than three surveys. .

Table A9.3: Frequency and direction of response bias in "Height" query: 2001, 2005 and 2009 HBSC surveys

Country	Sex	Sex Sign	Age	Age Sign	Wealth	Wealth sign	Health	Health sign
Austria	2	+	2	+	…	…	2	-
Belgium	1	-	3	+	…	…	3	-
Czech Rep.	…	…	2	+	1	-	1	-
Canada	…	…	3	+	…	…	3	-
Denmark	3	+	3	+	…	…	1	-
Estonia	2	+	3	+	2	+	3	-
Finland	3	+	3	+	…	…	1	-
France	1	+	3	+	…	…	…	…
Germany	1	+	2	+	…	…	1	-
Greece	…	…	1	+	…	…	2	-
Hungary	…	…	2	+	…	…	1	-
Ireland	2	-	3	+	1	-	…	…
Israel	…	…	2	+	…	…	2	-
Latvia	2	+	2	+	2	-	…	…
Lithuania	3	+	3	+	3	-	1	-
Luxembourg	1	+	2	+	…	…	…	…
Netherlands	2	+	1	+	…	…	…	…
Norway	…	…	3	+	…	…	1	-
Poland	…	…	3	+	1	-	…	…
Portugal	3	+	…	…	…	…	1	-
Slovenia	1	+	3	+	…	…	…	…
Spain	1	+	2	+	2	mixed	1	-
Sweden	3	+	3	+	2	-	2	-
Switzerland	3	mixed	3	+	…	…	1	-
Turkey	2	-	3	+	…	…	1	-
United Kingdom	…	…	3	+	…	…	2	-
United States	…	…	3	+	…	…	1	-

Source: Author's calculations of HBSC survey data for survey waves 2001/02, 2005/06 and 2009/10.

Note: "…" no evidence of response bias. For sex sign, positive means higher male non-response, negative is higher female non-response. For age, positive signs mean older students are significantly more likely to respond. For health and wealth signs, positive means richer or healthier students are significantly more likely to not respond, negative signs mean that and poorer or unhealthier are significantly more likely to be non-respondents. The table presents the results of a binary logit model. Bars indicate the frequency of statistically significant relationships between the identified correlates and likelihood of response to the "Height" question across three waves of the HBSC survey. Statistical significance is defined as $p \leq 0.05$. Greece, Israel, Luxembourg, and Turkey each participated in fewer than three surveys.