

Please cite this paper as:

Richardson, D. and N. Ali (2014), "An Evaluation of International Surveys of Children", *OECD Social, Employment and Migration Working Papers*, No. 146, OECD Publishing, Paris.  
<http://dx.doi.org/10.1787/5jxzmjrqvntf-en>



OECD Social, Employment and Migration  
Working Papers No. 146

# An Evaluation of International Surveys of Children

Dominic Richardson, Nabil Ali

JEL Classification: C83, J13

**DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS  
EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS COMMITTEE**

Cancels & replaces the same document of 17 February 2014

**OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS No. 146**

**An Evaluation of International Surveys of Children**

**Dominic Richardson and Nabil Ali**

*JEL Classification: J13 (Children), C83 (Survey Methods, Sampling Methods).*

*Key words: Child well-being, Child well-being data and indicators, Child well-being surveys.*

*Authorised for publication by Stefano Scarpetta, Director, Directorate for Employment, Labour and Social Affairs.*

All Social, Employment and Migration Working Papers are now available through the OECD website at [www.oecd.org/els/workingpapers](http://www.oecd.org/els/workingpapers)

Complete document available on OLIS in its original format

*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.*

**DIRECTORATE FOR EMPLOYMENT, LABOUR AND SOCIAL AFFAIRS**

[www.oecd.org/els](http://www.oecd.org/els)

**OECD SOCIAL, EMPLOYMENT AND MIGRATION  
WORKING PAPERS**

[www.oecd.org/els/workingpapers](http://www.oecd.org/els/workingpapers)

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to the Directorate for Employment, Labour and Social Affairs OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This series is designed to make available to a wider readership selected labour market, social policy and migration studies prepared for use within the OECD. Authorship is usually collective, but principal writers are named. The papers are generally available only in their original language – English or French – with a summary in the other.

*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*

**Applications for permission to reproduce or translate  
all or part of this material should be made to:**

**Head of Publications Service  
OECD  
2, rue André-Pascal  
75775 Paris, CEDEX 16  
France**

**Copyright OECD 2014**

## **ACKNOWLEDGEMENTS**

The authors would like to thank the coordinators of the surveys discussed in this paper and Willem Adema, Francesca Borgonovi, Monika Queisser and Conal Smith for helpful discussions and valuable comments on an earlier draft of the paper. Thanks also to Sophie Bournot and Pauliina Patana for statistical support.

## SUMMARY

This report evaluates sources of international child well-being data to assess their suitability for supplementing national and transnational data sources to inform policy. The review of the leading surveys of children (and surveys of households with children) summarises the information available from these sources and, as importantly, identifies the gaps in measuring child well-being outcomes not covered by data from these sources. The report then undertakes an in-depth evaluation of possible systematic bias in the underlying survey population to provide confidence in the reliability of outcomes measured from these international surveys. Based on the overall evaluation, the report concludes with recommendations for the use and improvement of international surveys for monitoring child well-being.

The report aims to answer the following key questions regarding current available sources of child well-being data: (i) What available data can be used to assess the well-being of children across countries? (ii) What are the strengths and weaknesses of existing international surveys for informing policy and monitoring the lives of children? (iii) How can the data for cross-national monitoring of child well-being be improved? In order to answer these questions three main activities were undertaken as part of the evaluation:

- A systematic review of the international data sources on children outlining the content of the survey datasets, methodology and organisational processes, and identifying the gaps in data within these sources.
- An evaluation of the strengths and weaknesses of the international data sources based on empirical testing of common indicators of child well-being derived from these sources, which provide an indication of the systematic bias that can exist in the underlying survey population and help inform comparability of outcomes calculated from the varying surveys.
- A presentation of key 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 to cater for bias in the underlying population when using these datasets for monitoring child well-being across countries.

## RESUMÉ

Ce rapport est une évaluation des sources de données internationales sur le bien-être des enfants. Il a pour objet de voir dans quelle mesure ces sources peuvent être utilisées en complément des sources nationales et transnationales pour éclairer les politiques publiques. Il commence par examiner les principales enquêtes sur les enfants (ainsi que sur les ménages avec enfants) afin d'inventorier les données qu'elles contiennent et surtout, de recenser les éléments du bien-être des enfants qui ne sont pas couverts par ces sources. Vient ensuite une évaluation approfondie des éventuels biais systématiques de la population observée dans ces enquêtes, ce qui permet d'évaluer la fiabilité des éléments mesurés dans ces études internationales. En conclusion, le rapport donne des recommandations, établies en fonction de l'évaluation globale, pour l'utilisation et l'amélioration des études internationales utilisées dans le cadre du suivi du bien-être des enfants.

L'objet de ce rapport est de répondre à certaines questions clés sur les sources de données actuellement disponibles sur le bien-être des enfants: (i) Quelles données disponibles peuvent être utilisées pour l'évaluation internationale du bien-être des enfants ? (ii) Quels sont les atouts et les faiblesses des enquêtes internationales existantes pour éclairer la formulation des politiques et connaître les conditions de vie des enfants ? (iii) Comment les données internationales de suivi du bien-être des enfants peuvent-elles être améliorées ? Pour répondre à ces questions, cette évaluation a consisté en trois activités principales :

- Un examen systématique des sources internationales de données sur les enfants, inventoriant le contenu des séries, expliquant la méthode suivie, détaillant les processus organisationnels et recensant les données manquantes dans les différentes sources.
- Une évaluation des atouts et des faiblesses des différentes sources internationales de données, grâce à un test empirique d'indicateurs courants du bien-être des enfants dérivés de ces sources, ce qui donne une idée du biais systématique de la population de chaque enquête et nous éclaire sur la comparabilité des résultats calculés à partir des différentes études.
- L'élaboration de recommandations clés pour l'utilisation et l'amélioration des données extraites des études internationales sur les enfants, indiquant la marche à suivre pour combler les lacunes existant, et la formulation de recommandations méthodologiques pour prendre en compte le biais de la population observée lorsque ces séries sont utilisées dans le cadre d'un suivi international du bien-être des enfants.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS AND DISCLAIMER .....	3
SUMMARY .....	4
RESUMÉ.....	5
CHAPTER 1. REVIEWING INTERNATIONAL DATA ON CHILDREN’S WELL-BEING: MOTIVATION, METHODS AND KEY RECOMMENDATIONS .....	9
1.1 Motivation: Why are child well-being comparisons necessary, and why is an evaluation needed?.....	9
1.2 Surveys to be reviewed: .....	10
1.3 Survey selection process: when, how, who and where? .....	12
1.4 How to evaluate the content of international surveys of children? An outline of the analytical framework.....	13
1.5 Key recommendations .....	14
CHAPTER 2. A SYSTEMATIC REVIEW OF INTERNATIONAL SURVEYS OF CHILDREN: SURVEY CONTENT AND GAPS IN CHILD WELL-BEING DATA.....	17
2.1 Systematic review of international surveys of children .....	17
2.2 Survey content and gaps in child well-being data.....	28
2.3 An overview of the child surveys, methods, data and gaps .....	41
CHAPTER 3. AN EMPIRICAL EVALUATION OF INTERNATIONAL CHILD SURVEYS: ARE THEY RELIABLE? .....	43
3.1 Why worry about bias in international surveys?.....	43
3.2 Analytical framework for assessing forms of bias .....	44
3.3 Analysis of bias and examples from the literature .....	50
3.4 Comparisons of interchangeable measures: external and internal reliability.....	69
3.5 Summary discussion .....	79
CHAPTER 4. RECOMMENDATIONS FOR THE USE AND IMPROVEMENT OF INTERNATIONAL DATASETS ON CHILDREN.....	80
4.1 How can the existing data for cross-national monitoring of child well-being be improved? .....	80
4.2 The way forward on three key issues.....	86
4.3 Which indicators can be used, now, for monitoring child well-being .....	91
4.3 Summary .....	95
ANNEX .....	96
REFERENCES .....	97

## Tables

Table 1.1: Countries participating in the six selected child surveys .....	11
Table 1.2: Criteria for child survey selection.....	13
Table 2.1: Structure of the systematic review of child surveys .....	17
Table 2.2: Survey country coverage and changes in the past three waves .....	22
Table 2.3: A summary of socio-demographic indicators .....	30
Table 2.4: A summary of education and schooling indicators.....	31
Table 2.5: A summary of health and risks indicators .....	34
Table 2.6: A summary of income and deprivation indicators .....	36
Table 2.7: A summary of housing and environment indicators .....	37
Table 2.8: A summary of civic participation and time use indicators .....	38
Table 2.9: A summary of relationship indicators.....	39
Table 2.10: A summary of subjective well-being indicators .....	41
Table 2.11: Summary table of cross-national child well-being indicators by dimension in the EU / OECD regions.....	42
Table 3.1: Analytical framework for empirical exploration of various forms of bias .....	45
Table 3.2: An overview of indicators used in comparative child well-being analysis.....	47
Table 3.3: The most commonly used survey-based child well-being indicators .....	50
Table 3.4: After sensitivity testing, enrolment rate associations weaken and most significant dropout rate associations disappear .....	52
Table 3.5: Strong associations between enrolment and dropout rates do not hold for high enrolment countries.....	53
Table 3.6: Higher participation rates are consistently associated to lower error and lower deprivation, but a higher range of aspirations .....	57
Table 3.7: Outside of reading literacy error margins there is little evidence to suggest systematic sample bias in the child well-being results.....	58
Table 3.8: Significant differences in means and variances between missing and non-missing values in selected PISA 2009 indicators .....	60
Table 3.9: Most countries have seen small non-response rates for parents' migrant status since 2000 ....	69
Table 3.10: Percentage of youth population in each survey reporting low affluence by type of scale.....	71
Table 4.1: An example of adjustment weights by sex and family affluence .....	88
Table 4.2: Suggested socio-demographic items for harmonisation across surveys .....	90
Table 4.3: Suggested socio-economic items for harmonisation across surveys .....	91
Table 4.4: Suggested items for complementing Social Protection Committee monitoring recommendations .....	93

## Figures

Figure 2.1: Dates of waves for the various international child surveys, 1983 to 2011 .....	21
Figure 3.1: There is little evidence to suggest school starting ages systematically bias child well-being measures.....	54
Figure 3.2: Less able students are not completing important contextual information for analysis in PISA59	
Figure 3.3: There is no clear pattern of changing non-response rates in aspirations between the waves ..	62
Figure 3.4: Evidence suggests that the placement of complex items in background questions does not systematically affect the precision of responses on parental employment.....	63
Figure 3.5: In 2009 fewer German children provide useable responses to the cellular phone ownership question.....	68
Figure 3.6: Correlation between PISA FAS calculations and HBSC FAS calculations .....	71

Figure 3.7: Proportion of households with children is similar in ESS and EU SILC .....74  
Figure 3.8: Household income distribution in ESS and EU SILC by household type, 2008.....76  
Figure 3.9: Associations between smoking prevalence and cannabis use as reported in HBSC and  
ESPAD.....78

**Boxes**

Box 3.1: Definitions of bias in the study .....44  
Box 4.1: Harmonisation: Methods for the statistical matching of survey data .....84  
Box 4.2: The value-added (or not) of interviewing children directly .....86

## **CHAPTER 1. REVIEWING INTERNATIONAL DATA ON CHILDREN'S WELL-BEING: MOTIVATION, METHODS AND KEY RECOMMENDATIONS**

1. To assess whether existing surveys of children, and surveys of households with children, can supplement national and transnational sources used for monitoring living standards in advanced economies, some important 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 (persons aged 0-17 years)? Does the content match presently understood concepts of child well-being measurement, and how might the content quality be assessed?

2. This chapter provides the motivational background to this work (section 1.1), the surveys to be reviewed (section 1.2), the selection criteria for the chosen surveys (section 1.3) and an outline of the analytical framework for the review (section 1.4), before concluding with the key recommendations for improvements derived from this evaluation (section 1.5).<sup>1</sup>

### **1.1 Motivation: Why are child well-being comparisons necessary, and why is an evaluation needed?**

3. Child well-being is an important 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 recent years there has been 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), which have fuelled the ongoing public debate.

4. 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 being comprehensive in capturing the life experiences of children, being calculated in an arbitrary fashion (for example, use of equivalisations and thresholds in outcome measures), 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 outcomes.

5. The evidence to come out from these recent cross-national comparisons of child well-being indicators has added to the need to better understand the quality of the data, and what the data truly represents. Recent reports have shown that (i) well-being measures can evolve in different directions (such as infant mortality and low birth weights – see OECD, 2011); (ii) virtuous trade-offs between efficiency and equity outcomes are evident (Freeman et al., 2010); (iii) poverty, family affluence and family structure,

---

<sup>1</sup> 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.

do not associate strongly with many important child-well-being indicators; (iv) although public spending is associated to child well-being measures some countries are getting more for their money (OECD 2009 and 2011 and Förster and Richardson, 2011); and finally (v) some indicators are more malleable over the short term than others.

6. For these findings to feed into policy decisions, there is a need for confidence in the reported child well-being statistics. Data on children, and the quality of their lives, will feed in to public debate about how to spend public money in the context of several other important competing demands for public spending. If statistics are incorrect or misleading, limited public money might be wasted to the detriment of the competing demand on the public budget, and on societal development on the whole. Moreover misrepresenting the actual state of children's 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 miss the children they are designed to help, and erode public confidence in similar evidence base responses in the future, to the detriment of all children.

### **1.2 Surveys to be reviewed:**

7. Six child surveys were reviewed as part of this evaluation:

- 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).

8. Table 1.1 outlines country participation for the selected surveys in the most recent and upcoming waves. Across the OECD Japan and Mexico take part in only two of the six surveys reviewed. All but two (Cyprus<sup>2,3</sup> and Portugal) European Union (EU 27) countries take part in at least four of the six surveys The table also shows that 17 countries in total are members of all six surveys although, for certain survey waves Belgium is represented by only one community, and the United Kingdom does not include one or more of the home countries (England, Northern Ireland, Scotland or Wales). Each of these 17 countries is a member of the EU 27, and with the exception of Bulgaria, Latvia and Lithuania these countries are also members of the OECD.

---

<sup>2</sup> 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".

<sup>3</sup> 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.

Table 1.1: 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 <sup>1,2</sup>	✓	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 <sup>1</sup>	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

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.

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 United Kingdom is represented by English and Scottish samples only. PIRLS 2006 countries not in 2011 include Flemish speaking Belgium, Iceland, and Latvia. TIMSS 2007 countries not in 2011 include: Bulgaria, French speaking 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 speaking Belgium did not take part in ICCS 2009 and only England took part for the United Kingdom.

\* See footnotes 1 and 2 on page 10.

Sources: Survey websites, Accessed May 2011.

9. Additionally three household panel surveys were also included in this 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. These surveys do not ask questions directly to dependent children under 18 (but to head of households from age 15; and from age 16 for EU-SILC).

### 1.3 Survey selection process: when, how, who and where?

10. The child surveys included in this project meet four important 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 United Nations Convention on the Rights of the Child (UNCRC) age definition of childhood).** 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 available to 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 for inclusion. Surveys with insufficient numbers of EU or OECD countries included in international child well-being reports are not reviewed simply because they would not allow for substantive cross-national comparison. The low threshold (at least 50% of OECD and EU member states) for the selection of surveys is because of the potential that surveys have for expansion (inclusion of more countries) in the future.

11. Each of the child surveys listed in Table 1.1 have met these criteria. Table 1.2 below shows the coverage of children in each survey by age, sample type, and the unique content of the survey.

**Table 1.2: Criteria for child survey selection**

	<b>EU School Project on Alcohol and other Drugs (ESPAD)</b>	<b>Health Behaviour in School-aged Children Study (HBSC)</b>	<b>International Civic and Citizenship Education Study (ICCS)</b>	<b>Progress in International Reading Literacy Study (PIRLS)</b>	<b>Programme for International Student Assessment (PISA)</b>	<b>Trends in International Mathematics Science Study (TIMSS)</b>
<b>Most recent wave (next wave)</b>	2007, 2011, (2015)	2009/10, (20013/14)	2009 (2016)	2006, 2011, (2016)	2009, 2012, (2015)	2007, 2011, (2015)
<b>Approx. Child age</b>	16 year olds	11, 13 and 15 year olds	8 <sup>th</sup> grade students (above the average age of 13.5 years)	4 <sup>th</sup> grade students (aged around 9.5 years)	15 year olds	4 <sup>th</sup> and 8 <sup>th</sup> grade students (aged around 9.5 and 13.5 years)
<b>Content of focus</b>	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)
<b>Countries (EU/OECD)</b>	25/23	26/28	22/23	23/25	25/34	22/27

Source: OECD Secretariat calculations.

12. The surveys selected run in waves of three (PISA) four (ESPAD, HBSC, TIMSS) or five (PIRLS) years, most recently undertaken in 2009, 2010 or 2012. In each of the surveys at least 23 OECD countries and at least 22 EU 27 countries are included. The six child surveys collect data from children between the ages of approximately 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 focus on different well-being data at different ages, 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**

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

##### ***A systematic review of international surveys of children***

14. A key starting point for assessing the availability and quality of comparative data on children is to review the surveys from which the data is derived. The first part of Chapter 2 undertakes a systematic review of the six child surveys and three household surveys. Each part of the systematic review contributes to the subsequent work undertaken in this report evaluating the gaps and bias in the data. As part of the review background information and technical reports for each survey have been consulted. Chapter 2 discusses the results of the systematic review with a particular focus on how the choices made for each survey will impact on its ability to contribute to comparisons, and monitoring, of child well-being.

### ***Gaps in child coverage and the child well-being data***

15. 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, importantly, what is missing.

16. Thus, following the systematic review of the surveys, the second part of Chapter 2 reports the gaps in the coverage of child well-being indicators and provides vital information for future development of surveys designed to inform cross-national monitoring of children. This review of 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 (Chapter 4).

### ***Empirical evaluation of international datasets on children: are the resulting child well-being indicators reliable (what are the biases in the surveys)?***

17. 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.

18. The structure of Chapter 3 is organised by biases that are (i) system-based, (ii) sample or methodologically based, and (iii) culturally or linguistically based. This empirical evaluation of the comparability and validity of the indicators in the surveys also reviews examples in the literature of efforts to explore forms of bias in the nine surveys before performing tests on the data itself. Chapter 3 concludes by identifying methods which have the greatest success in reducing biases of these types. These methods will be identified for ‘good practice’ purposes and highlighted in the detailed recommendations provided in Chapter 4. Where biases are identified, implications for appropriate interpretation of the survey results have also been proposed.

## **1.5 Key recommendations**

19. 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 allow for timely comparisons, as well as comparisons of trends over time and between dimensions; it would represent the overarching latent concept of child well-being, but also easily break down to dimension and indicators-level in order to encourage considered policy / civil society responses. It is important to know what is needed in order to achieve this ‘gold standard’ monitoring tool.

20. 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 Chapter 4 and the analysis undertaken in Chapters 2 and 3:

- **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 are available, survey data are needed for detailed analysis and recommendations. Some survey data for European countries are already available for this purpose (for example, data on mental health from the ESPAD survey), but 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 from wave to wave, and consult users (through open forums or bilateral discussions). Policymakers and researchers should be encouraged to provide evidence in support of claims to the changes they request, 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 equality 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 survey items.
- **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 make *post hoc* weighting adjustments to indicators with bias in non-response; when more complex multivariate analysis is undertaken researchers should ensure representative estimates are used to fill gaps. 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.** To facilitate this, in the first instance, survey coordinators should report on missing populations (with the help of national survey managers), or consider supplemental out-of school surveys. Exclusion can occur for various reasons, though most often active exclusion (students actively deselected from the sampling frame) is due to perceptions regarding student being '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. Such exclusions should be clearly stated (as in ICCS) and more efforts should be made for future inclusion of these groups as they often include vulnerable children that require special policy interventions.
- **Efforts should be made to harmonise data collections between the major international surveys.** This will facilitate analysis of 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 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.
- **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 2. A SYSTEMATIC REVIEW OF INTERNATIONAL SURVEYS OF CHILDREN: SURVEY CONTENT AND GAPS IN CHILD WELL-BEING DATA

21. This chapter presents a systematic review of the six child surveys and three household surveys selected for evaluation (see Chapter 1), followed by a review of the existing gaps in data within these surveys. The aim of the systematic 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 this report (see Chapter 3). Identifying gaps in the data serves two purposes: first, gaps in the coverage of well-being indicators provides vital information for future development of surveys designed to inform cross-national monitoring of children; second, identifying gaps in the data allows for better understanding of how to appropriately interpret the available data used for monitoring the well-being of children. Results of the systematic review and the gaps analysis inform recommendations for good practice in, and set expectations for, the collection of comparable child data in the future (Chapter 4).

### 2.1 Systematic review of international surveys of children

22. Information on each survey's management, content, and collection methods has been organised under the following themes: (i) management structure, (ii) data set basic information, (iii) methodology, (iv) delivery in the field, and (v) next wave/project. Under each theme, three to six subsections categorise the detailed information to be compared and assessed across the surveys. Table 2.1 outlines the structure of the survey review.

**Table 2.1: Structure of the systematic review of child surveys**

<b>Management Structure</b>	Statement of purpose Funding and affiliations Co-ordinators and expert panels Government partners
<b>Data set basic information</b>	Survey dates Countries covered and target populations Contents (including core and contextual information)
<b>Methodology</b>	Unit of analysis Questionnaire development The sample design The sample threshold and treatment of under-sampling The timing of the data collection window
<b>Delivery in the field</b>	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
<b>Next wave/project</b>	Next date and availability New content / questions if applicable Previous content not repeated in the next wave Countries added to, or leaving survey

23. For the systematic review of child surveys the data, questionnaires, background information, and technical reports for each survey have been consulted. For more detailed information on the structure of the surveys included in this review, please refer to Annexes<sup>4</sup> 2, 3 and 4 of this document.

### **2.1.1 Management Structure**

#### *2.1.1.1 Statement of purpose*

24. 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.

25. The three education surveys (PIRLS, PISA, and TIMSS) measure cognitive skills, yet differ slightly in terms of how (life skills or curricula based, reading, mathematics or science), for who (by age or grade), and the contexts (background data) by which the variation in cognitive skills might be explained. 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 trends in alcohol and drug habits among students aged 16, and to compare trends between countries and between groups of countries. 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 persons aged 11, 13 and 15 at the national and international levels, and to advance scientific knowledge’. 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.

26. 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 child surveys. Therefore, the expectations for the types of child well-being data derived from these surveys may be different from child-respondent surveys.

27. The most common stated purpose for each of the nine surveys is informing policy. Of all of these, only ESPAD and ICCS do not explicitly state informing policy as a key goal. However, well-being policies explicitly, are a low priority: of all 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, five of the surveys explicitly mention trends analysis as a key purpose (child surveys PIRLS TIMMS, and household surveys ESPAD, ESS and EU-SILC), although HBSC has recently completed three identical waves in order to be able to analyse trends but it is not a core design feature. Also, it is important to note that only two surveys explicitly mention the ‘comparison of distributions’ as a key purpose (TIMSS and ESS), despite the fact that nationally representative stratified micro-data available in many of the surveys allow for this type of analysis.

#### *2.1.1.2 Funding bodies and affiliated organisations*

28. Coordinating an international survey can be an 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 can place restrictions on survey coordinators when developing the surveys, or changing (or retaining) the structure or the content of the survey. Funders and affiliates represent the main

---

<sup>4</sup> The annexes in this document were prepared in the early stages of the project in late 2010. Before the paper will be released these annexes and associated information in the main documentation, will be updated to the most recent surveys waves, and checked with coordinators of the surveys.

stakeholders but do not operate the surveys, and can be transnational or national bodies, academic institutions, non-governmental organisations or government ministries.

29. 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, with the remaining costs covered by non-OECD countries also taking part in the survey.

30. 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 30 000 to 40 000 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.

31. A number of the European surveys receive support from European organizations. Other regional organisations also contribute to the costs of the other surveys or help with the dissemination of surveys. For example, the WHO Regional Office for Europe contributes to research dissemination and the publication of international HBSC reports.

32. 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.

33. 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); PISA (ministries of participating countries); EQLS (European Commission); ESS (European Commission); ESPAD (Swedish Ministry of Health); and EU SILC (the European Commission and the National Statistical Offices of EU member states).

34. 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. Support in-kind for EQLS, and involvement in the decisions on the direction and delivery of the survey, comes from independent research groups and universities in Germany, Italy, the Netherlands and Poland.

#### *2.1.1.3 Co-ordinators and expert panels*

35. 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 as they decide which items will be included or dropped from the questionnaires, topics for rotating modules, and which countries are accepted in future waves (as well as which countries might be encouraged to join). In cases where questionnaire development groups as well as management consortium exist, recommendations, contacts and questions regarding developments to specific parts to the surveys can be delivered directly to these groups.

36. ESPAD has a questionnaire group for developments and amendments. In recent waves the questionnaire groups have been informed by thematic groups. 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), with support of an elected Coordinating Committee and Assembly of Principal Investigators from each country. A joint management committee oversaw 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 International Association for the Evaluation of Educational Achievement (IEA). PIRLS and TIMSS is coordinated by the IEA International Study Centre (ISC) located at Boston College, United States. PISA assessments are developed co-operatively, agreed by participating countries, and implemented by national organisations.

37. The interview questionnaire for EQLS is developed by a research consortium consisting of academics from universities across Europe (including Bulgaria, Cyprus<sup>5</sup>, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, the Slovak Republic, Slovenia, and Turkey). ESS questionnaire development is headed by the Centre for Comparative Social Surveys, City University, UK. While the EU SILC questionnaire development is coordinated by Eurostat, with the actual development undertaken by an expert group, made up of national delegations of experts.

### ***2.1.2 Data set basic information***

38. 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.

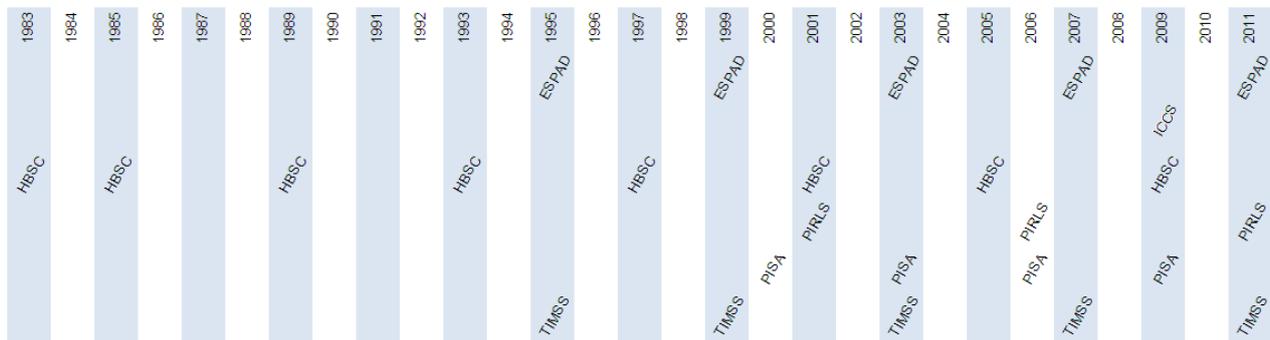
#### ***2.1.2.1 Survey dates***

39. The timing between updates (waves) of the data has an important bearing on whether or not the information collected by each of these surveys can be used for reporting or monitoring child well-being. 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 (Chapter 4).

40. 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 2.1). 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 repeated items), for developing trends or pooled time-series analysis, can also be derived from this pattern; with the exception of ICCS each of the surveys have at least three completed waves to date. The most robust trends by country (as measured by consistent repeated items and country coverage) across waves can be derived from ESPAD, HBSC, and TIMSS. Finally, it is also useful to know the dates when the surveys were undertaken previously in order for the data to be applied for use in natural policy experiments (that is, measuring change in outcomes either side of a policy change).

---

<sup>5</sup> See footnotes 1 and 2 on page 10.

**Figure 2.1: Dates of waves for the various international child surveys, 1983 to 2011**

Source: Selected child survey reports. See Annex 2.

41. 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 as they will coincide in future years, which suggests that efforts to harmonise surveys could concentrate, in particular, on these two studies.

#### 2.1.2.2 Countries coverage

42. The countries covered in each survey will inevitably affect the usability of the survey for cross-national comparisons in Europe and across the OECD. The coverage of countries in the most recent wave was an important criterion for the inclusion of surveys in this project (Chapter 1). Country coverage can also change between the waves, which will affect the possibility to construct time series.

43. Table 2.2 reports changes to country membership in the last three waves of each survey and projections for the upcoming wave. There are two important patterns. First, all of the surveys are increasing in size over the years, which means that demands on schools in particular will be increasing, and efforts to innovate collection methods will be necessary over time (for an example see PrePIRLS at <http://timssandpirls.bc.edu/pirls2011/prepirls.html>). Second, a handful of countries are not permanent members of the surveys, for example the OECD countries Greece, Israel and Turkey enter and leave the surveys. Also, some developed countries have never been part of unique health and risk surveys (ESPAD and HBSC), such as Australia, Chile, Japan, Korea, Mexico and New Zealand.

**Table 2.2: Survey country coverage and changes in the past three waves**

	Countries covered <sup>6</sup>	Countries not repeating next wave	Countries to be added
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, the 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 Behaviour in School-aged Children	2000/01, 36 countries including: Austria, Belgium, 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	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, the 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 including: Bulgaria, Canada (Ontario, Quebec), Cyprus, the Czech Republic, France, Germany, Greece, Hungary, Iceland, Israel, Italy, Latvia, Lithuania, the Netherlands, New Zealand, Norway, Romania, the Russian Federation, the Slovak Republic, Slovenia, Sweden, Turkey, the United Kingdom (England and Scotland), and the United States. PIRLS 2006, 40 countries: including additionally Belgium, 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, (the 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 including: Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the 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, 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, 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, Italy, Japan, Korea, Latvia, Lithuania, the Netherlands, New Zealand, Norway, Romania, the Russian Federation, the Slovak Republic, Slovenia, Spain (Basque Country), Sweden, the United Kingdom (England, Scotland), and the United States (including Indiana State as a	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

<sup>6</sup> For Cyprus, see footnotes 1 and 2 on page 10.

	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).		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, Slovenia, Spain, Sweden, Turkey and the United Kingdom. In 2007, 29 Countries: as above and 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 including: 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 and Bulgaria, Cyprus, Latvia, Romania and the Russian Federation.	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.

Notes: Countries listed in this table only include countries of focus covered in this study (OECD; EU member states). Full lists of countries covered by the surveys are available in the annex of this document. Bulgaria is still a member of HBSC, but was unable to complete the survey in 2009/10 due to problems with funding.

Source: Survey websites, accessed May 2011.

### 2.1.2.3 Survey contents

44. Understanding the content of the surveys is important for assessing the extent to which they can inform comparisons 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 used 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. In this section we look at the broad content of the surveys and not the content by item, which is covered in detail in the gaps analysis (section 2.2) and in the Annexes 2, 3 and 4 of this report.

45. Core and contextual questionnaires are primarily differentiated in educational or civic studies that include a test questionnaire and background questionnaire (“two-step collection method”). 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 and background questionnaires in the surveys is to record the aspects of a child’s family background that might help to explain variation in students’ achievement.

46. ICCS, for example, describes a two-step collection method with a core ‘assessment’ questionnaire covering a civic and citizenship framework with cognitive tests, which is supplemented by a ‘student’s perceptions’ questionnaire which provides background information on socio-demographics of students and their families, classroom environments, and teaching and learning, and provides the ‘contextual framework’. PIRLS takes a similar approach to ICCS, by including a core reading assessment

based on multiple choice and written responses (including reading for literary experience and reading to acquire and use information), and background questionnaires. The approach taken by TIMSS is similar to PIRLS, with the exception that the core questionnaire assesses students in mathematics and science. While 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. PISA also uses 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).

47. 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 alongside socio-demographics in the ESPAD questionnaire are questions on alcohol, tobacco and drug use, leisure time use, educational attainment, and also questions 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, injuries), and health outcomes (symptoms, life satisfaction, self-reported health, Body Mass Index)<sup>7</sup>.

48. Household surveys tend to consist of a core questionnaire and rotating modules that look at specific issues just once (or on a longer cycle, such as every four years). The core questionnaire of each survey includes socio-demographic information such as household size, employment status, level of education and marital status. However, the questions in the survey used to collect this data can change in terms of wording and scaling.

49. 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. More detailed information is collected by the ESS questionnaire as it includes two sections of similar sizes in the main questionnaire (approximately 120 items); a 'core' module and two or more 'rotating' modules. A supplementary questionnaire is also 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.

50. EU-SILC collections focus on income but also cover information on housing, labour, health, demography, and 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.

### **2.1.3 Survey Methodology**

#### *2.1.3.1 Unit of analysis*

51. The levels at which analysis can be undertaken on survey data are the unit of analysis. In complex surveys, such as 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 the flexibility of the surveys for analysis of child well-being outcomes, and can also influence recommendations for best practices in weighting and appropriate analysis. Each of the child

---

<sup>7</sup> For detailed information on optional modules or questionnaires in each survey please refer to Annex 2.

surveys included in this report are undertaken in a school setting, although the sample methods for each survey varies.

52. Both of the health and risk surveys report results primarily at the student level. ESPAD surveys students who 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, for example, children with special educational needs or who are severely handicapped, are not included in the ESPAD target population. In ICCS, eighth grade students in participating countries are principal units of analysis. Units of analysis also include schools and educational systems. 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 as well. 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 steps, the first of which samples schools (stratified proportionate to the size of the school). In both PIRLS and PISA, primary care givers also receive questionnaires as part of the collection.

53. 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 and includes all persons aged 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 (Denmark, Finland, Iceland, the Netherlands, Norway, Sweden and Slovenia), only a selected household respondent receives a personal questionnaire and household and income variables are collected either via register or the selected respondent. More information on the target population is available in Annexes 2 and 3.

#### *2.1.3.2 The timing of the data collection window*

54. 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.

55. When the data collection 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.

56. In the case of most global 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 from October to December the year before the survey year (so for PIRLS 2006, southern hemisphere countries were surveyed in 2005), and from March to June of the survey year for northern hemisphere countries. A similar collection window is used for ICCS 2009 (the survey took place from October to December 2008 in the southern hemisphere and from February to May 2009 in the northern

hemisphere), and PISA where testing is conducted in a 42 day period between 1 March and 31 August. 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 fieldwork for each cross-national survey is carried out over a period of around seven to eight months, from October to May.

57. Household surveys, also have 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 the survey year. EU-SILC has a similar window and 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.

### *2.1.3.3 The sample design*

58. A final, and important, aspect of survey methodology is the sampling methodologies used. The design of samples to achieve the best representativeness of the target population changes from survey to survey, and can affect the treatment (weighting; significance tests) required of final results to produce accurate estimates of population experiences across child well-being measures. In recent years it has become clear that more and more surveys of children have put a great burden on schools. 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 thus affecting overall comparability.

59. 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. See Annexes 2 and 3 for more details on sampling methodology.

60. The goal of the sampling process in ESPAD is to obtain a national and "gender-wise" representative dataset, but sampling methods are left up to each national researcher. The sampling strategy of HBSC aims to achieve nationally representative groups at the mean ages of 11.5, 13.5 and 15.5 and the sampling frame (the identified population from which the respondents will be sampled) aims to include 95% of the target population (the minimum recommended sample size is 1500 children per age group, and thus, 4500 children per country). PISA sampling undertakes a two-stage probability proportionate to size. 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. In the second stage, students are sampled in schools randomly until 35 students are selected with equal probability (all 15-year-old students are selected if fewer than 35 are enrolled). For TIMSS and PIRLS, as with PISA, probability proportionate to size is used to select the schools. Unlike PISA, within school sampling randomly selects a class group (cluster sampling). For ICCS, again, probability proportionate to size 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.

61. For EQLS 2003 and 2007 a multi-stage stratified random sample was used which draw sample points from stratification by region and degree of urbanisation (random route method is used for clusters of 20 houses after the initial house is drawn at random). The EU-SILC cross-sectional<sup>8</sup> data collected (whether survey or register based), are based on a nationally representative probability sample of the population residing in private households within the country, irrespective of language and nationality.

62. 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

<sup>8</sup> Information on longitudinal sampling is not included here (see Annex 2).

in all of test-based studies). What is not completely clear are the reasons behind the exclusion of school and pupils from samples in all surveys. Exclusion can occur for various reasons, though most often active exclusion (students actively deselected from the sampling frame) is 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. Also the definition of such students can also vary between countries and may add bias in the underlying sample (see Chapter 3). Please refer to Annexes 2 and 3 for more details on sample size thresholds for each of the surveys.

#### ***2.1.4 Delivery in the field***

63. How the training and management of field staff is undertaken by the survey coordination team is likely to affect the quality of the surveys. HBSC is run on a network model, involving experts from across all participating countries. 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. 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. Boston College serves as the International Study Centre for IEA's studies in mathematics, science, and reading - TIMSS and PIRLS.

64. Household surveys 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, and carried out the data collection in each country. By contrast, the ESS relied on the Central Coordination Team (CCT) to design and coordinate the survey. Central coordination entails participation from all responding countries, and independent advice from teams of independent academics.

65. Good quality translation is also an integral part of ensuring that respondents answer the same questions across countries, and that the responses are delivered without bias. Each international survey has translation procedures, although these vary in terms of the complexity and the management approach. For different surveys translation methods can take the form of either recommendations or requirements. For example, for countries joining ESPAD a translation and 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 'through double translation and reconciliation' (PISA technical report, 2009). In household surveys, examples of the management of the translation process include decentralised and outsourced, and centralised processes. For instance, 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.

#### ***2.1.5 Next wave/project***

66. One of the key criticisms of the attempts to monitor child well-being cross-nationally has been timeliness. The dates for the next child surveys are outlined in Annex 2. The latest collections of HBSC and PISA were completed in 2012, 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 study's 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. 2007 data was only available via an online request form. TIMSS and PIRLS, both 2011, were also not yet available.

67. 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 were 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). 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 two years after the collection, in March (provisional) and August (final) respectively. Finally, ESS 2006/07 was made available to registered researchers in September 2009. Details of the changes in questionnaire content between waves are available in Annex 2 and 3.

## 2.2 Survey content and gaps in child well-being data

68. In order to assess the contents of international surveys for measuring child well-being it is as important to look at what is missing – the “gap” – alongside what data are included. This section provides an overview of the indicators available from the selected surveys, and highlights the gaps in the data that need to be filled for more comprehensive monitoring of child well-being outcomes.

69. A review of the indicators available background questionnaires<sup>9</sup> from cross-national surveys of children was undertaken for each survey and across each wave, followed by a systematic analysis to identify how these responses are collected and translated into data points in the survey datasets. This data is then used to categorise the information into dimension, and further, into domains within dimensions. Finally a frequency analysis is used to provide a summary of each dimension. Dimensions are defined after collating all questionnaire items into a database (see Annex 4), and categorising these items into over-arching policy groups, socio-demographic items, relationship items and subjective perceptions of well-being. Items with dimensions are then re-ordered into commonly recognised domains.

70. Given the objectives of this report it is logical to first look at 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).

71. 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. In the frequency analysis the number of scale, ordinal, nominal and binary indicators are compared for the future purposes of assessing the method with which child well-being data is collected, and recommending appropriate forms of analysis for that data. The tables in each section below summarise the number and types of indicators, under each dimension, included in the selected surveys (Tables 2.3 – 2.10).

---

<sup>9</sup> This review did not include non-compulsory questionnaires, such as the PISA parental questionnaire.

### 2.2.1 Socio-demographics dimensions

72. This section reviews the socio-demographic data available and identifiable gaps from the surveys reviewed, organised into seven domains of age, family form, languages, migrant status, parental education, religion and sex. 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, etc.).

- 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 an adult (usually the head of household), and can be reported as completed year or year/month units. Child age in months or weeks is not 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). HBSC further asks children whether they live in a foster home or a children's home.
- **Family form** in 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).
- Items recording the **migrant status** of the child or family are a regular addition to the surveys, with the exception of HBSC, where this is included in the optional package (undertaken by around one-third of countries in 2009/10). The question is asked in different ways across the surveys, the simplest of which is found in ESPAD, put as 'Where were you born?' The household 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 born outside the country 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. 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 and EU-SILC asks respondents for their education level, and also their partners' and their parents' education level. 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 been included only in one cross-national survey in the OECD and European region in recent years. The ESS 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.

**Table 2.3: 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

### 2.2.2 Policy dimensions

73. This section reviews the objective (and non socio-demographic) items included in each of the surveys and organises the data in policy dimensions. Each dimension contains information organised by domain before these results are summarised in a table.

74. It should be noted that this section does not attempt to systematically discuss which concepts are missing by dimension, as this would require a judgement to be made on the use of presently collected items to serve as proxies, or to have an opinion of the relative value of additional inclusion over existing measures (neither of which is fully testable without access to an international child well-being survey). Rather, recommendations for the collection of new items, or adjustments to the methods of collection of present items, is addressed in section 4.

#### 2.2.2.1 Education and schooling

75. The dimension 'education and schooling' has been organised into the domains of: general education, attainment and achievement, attendance, early years attendance, extracurricular activities, homework, and teaching and learning. Selection of domains in the education and schooling section were made on the basis of the 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.

- **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 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 the student started primary school.
- **Extracurricular activities** are included in the PISA background questionnaire and ask students about their attendance of extension courses, remedial courses 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 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 (there is 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 practices. 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 whether teachers represent all sides of an argument as part of teaching. PIRLS asks two questions about teaching practices, 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).

**Table 2.4: 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

### 2.2.2.2 Health and risks

76. The domain ‘health and risks’ has been sub-categorised into: 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. Selection of domains in the health and risks dimensions follow commonly applied 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]), behaviours 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 assessment of their general health (different from subjective assessment of feeling).<sup>10</sup> 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**.<sup>11</sup> The household surveys all ask questions on the subject. ESS and EU SILC ask about long standing illness and 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.
- 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 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

<sup>10</sup>. It is important to distinguish between self-reported and subjective measures. This report considered measureable self-reported factors not to be subjective when they consider counts or unbiased information, such as a respondent's self-reported age.

<sup>11</sup> 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).

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 having 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. **Sex** as a risk behaviour is also 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. 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 ('relewin' 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).
- 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.
- 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.

**Table 2.5: 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

### 2.2.2.3 Income and deprivation

77. The domain ‘income and deprivation’ has been organised into the sub-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.

- 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, and 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** items in the surveys reflect on whether households can make ends meet on their present incomes. ESS asks for an objective assessment of going without (rather than income required as above). EU-SILC asks about 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. 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 the respondent goes to school or 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 2009 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.

**Table 2.6: 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

#### 2.2.2.4 Housing and environment

78. The dimension ‘housing and environment’ has been organised into the domains of: main home, housing conditions, housing security, tenure, locality, crime, and environment. Housing and environment domains are selected to separate out immediate environments and conditions (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 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 asks 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, and if they experience pollution or environmental problems.

**Table 2.7: 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

### 2.2.2.5 Civic participation and time use

79. 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. 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).

- Questions 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, sign petitions, how interested they 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 or a trade union, and plans to stand in elections when they are old enough.
- **Social participation** questions are asked in ICCS directly to children, and in EQLS and ESS to household respondents. The ICCS question is simply put as ‘How interested are you in social issues?’ 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 their respondents how many hours a week they spend watching TV, listening to the radio, reading newspapers and surfing the internet. HBSC include items on time watching television and playing computer games.
- **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 in the 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. HBSC asks questions about the **time spent with friends** 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 EQLS, where adult respondents are asked how often they attend religious services (excluding weddings and funerals and other main religious events).

**Table 2.8: 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

#### 2.2.2.6 Indicators of personal relationships

80. Personal relationships, a critical aspect of personal well-being, and can be influenced by policy choices such as anti-bullying programs and parental classes. In the conceptual frame of well-being, relationship indicators can also be treated as contextual variables – the quality of which can add to or diminish value from policy inputs to achieve other policy goals (such as educational achievement). The relationships dimension is organised into the domains of: family and social relationships, school relationships, and antisocial behaviour. Selection of domains in this dimension 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 adults within the household 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). HBSC includes a question 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 receive 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<sup>12</sup>. 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, stole 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.

**Table 2.9: 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; 15/16	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

### 2.2.3 Indicators of subjective well-being and opinions

81. Indicators of subjective well-being are presented separately as personal reflections on levels well-being. This includes measures of satisfaction and aspiration, and is commonly separated from objective perceptions of well-being, with anchored references, self-reported or otherwise (OECD, 2009; OECD, 2013; and 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. The selection of domains is based on subjective perceptions of indicators that, objectively collected, would be placed in the dimensions and domains outlined above.

<sup>12</sup>

In 2009, these questions were included in the parental questionnaire.

- **Aspirations** refer to the child's perception of their future. There are 3 questions in total, asked of children aged 13 and 15 in TIMSS and PISA respectively. TIMSS asks 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 asked specific questions in 2006 about how teaching of science in school might prepare the child for a science career, 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, and how teachers rate the child's 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 used by HBSC which measures life satisfaction based on the Cantril'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 covered in the ESPAD and ESS. 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.
- Information on **subjective material situation** 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.

**Table 2.10: 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 15; 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

### 2.3 An overview of the child surveys, methods, data and gaps

82. The systematic review<sup>13</sup> shows 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. 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 greater engagement with the users of the data for the benefit of consistency across the waves on key child well-being indicators.

83. Together the review and the gap analysis<sup>14</sup> shows that 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.

84. 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 2.11). 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. The next most populated dimension is income and deprivation, then health and risk

<sup>13</sup> Annex 2 and 3 provide detailed information on the structure of the surveys included in the systematic review.

<sup>14</sup> Annex 4 provides detailed information on existing and missing dimensions and domains of child well-being indicators.

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 up. In the former, adults' data are most often available, and in the latter, child data are the main source.

**Table 2.11: 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
No. 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

85. In terms of age, in each dimension data are missing for children under the age of 9 years. By dimensions, health and risks and housing and environment do not have age specific responses for children under the age of 11. Further, the data cannot effectively distinguish disabled children, children out of home, and in the case of HBSC, migrant children.

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

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

87. The aim of international surveys of children is for 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 forms of bias in the responses (Box 3.1). Bias in responses refer to differences between recorded values and true values that can occur because of the exclusion of certain respondents from a sample, non-response, and incomparable responses due to cultural or social endogeneity in the questionnaire items.

88. This chapter tests for bias in child survey data used in the construction of child well-being indicators. However, to test for all forms of bias across all items within the nine selected surveys is beyond the scope of this paper. Instead the analysis will focus on a range of tests undertaken using different surveys to demonstrate the various forms of bias that can exist in survey data during the creation of child well-being indicators.

89. 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, the analysis only covers data used to develop indicators for published comparative studies of child well-being. 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).

#### **3.1 Why worry about bias in international surveys?**

90. 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 and the reliability of trends for that country over time. With replicated bias, although the validity of the results are in question, comparability and reliability can be ensured if each country compared suffers from the same degree of inaccuracy in comparable waves.

91. 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 3.1: Definitions of bias in the study

Analysis in this paper is organised in terms of bias tests. Taken together the bias tests will feed into broad assessments of practice and will feed into the recommendations to be outlined in the final 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 ‘true’ 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. These types of government policies should be treated as contributing factors to variation in true values (such as policies 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 (Annexes 2 and 3). 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.

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.

### 3.2 Analytical framework for assessing forms of bias

92. The analytical framework that will guide the empirical evaluation of the data from each survey is outlined in Table 3.1. Tests will be presented in the order shown, and include a description of the indicator. Only summary results will be presented in the main text; detailed results are available in Annexes 5 – 8.

**Table 3.1: Analytical framework for empirical exploration of various forms of bias**

<b>Bias</b>	<b>Indicators</b>	<b>Tests</b>
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 Secretariat calculations.

93. Each analytical section will record significances associated with 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.

94. 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 on 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 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 by such analysis (Hanuschek and Woessmann, 2010). Efforts to differentiate between these forms of bias will be made in all parts of the analysis. However, the purpose of this chapter is only to inform indicator development and trends.
- Much of the previous research on cultural and linguistic bias 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.

### **3.2.1 Indicators selected for analysis**

95. Because each survey in this review has many items that could be used for child well-being indicator development, 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 to create child well-being indicators in recent comparative projects provides the foundation for selection.

96. Table 3.2 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, countries formerly in the Soviet Union). With the exception of the US Department of Commerce comparison from 1990, all these studies have been undertaken in the past seven years.

Table 3.2: An overview of indicators used in comparative child well-being analysis

	European multidimensional indices (Bradshaw et al. 2007 and 2009) <sup>1</sup>	UNICEF Report Cards 7 (2007) and 9 (2010) <sup>1</sup>	CEE CIS index (Richardson et al., 2008)	Doing Better for Children (OECD, 2009)	Save the children (2008 and 2011) <sup>1</sup>	European Commission Child poverty and Well-being (2008) <sup>2</sup>	European Commission TARKI (2011)	US Dept. of Commerce (Hobbs and Lippman, 1990) <sup>3</sup>	OECD Family Database (since 2007)
<b>Health and safety</b>	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 <sup>c</sup> ; life expectancy <sup>a</sup> ; infant mortality <sup>b</sup> ; low birth weight <sup>b</sup> ). Health behaviour (nutrition <sup>b</sup> - breakfast; nutrition <sup>c</sup> - 5 a day; dental <sup>c</sup> x 2; physical activity x 2 <sup>b,c</sup> ). Health Problems (BMI <sup>b</sup> and <sup>c</sup> ; mental health <sup>c</sup> ; morbidity - chronic <sup>c</sup> ; morbidity - infectious <sup>c</sup> ). Mortality rates (suicide <sup>b</sup> ; suicide and self-harm 10-24 <sup>c</sup> ). Vaccinations <sup>c</sup>	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
<b>Subjective well-being</b>	Subjective well-being (life; peers x 3 [2007]; health).	Subjective well-being (health; life; peers x 3)				Subjective well-being (home <sup>c</sup> ; school safety <sup>c</sup> x 2)	Subjective well-being (life)		
<b>Personal family and peer relationships</b>	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 <sup>b,c</sup>			
<b>Material well-being and economic security</b>	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 <sup>a</sup> ; relative gap <sup>a</sup> ; persistent <sup>b</sup> ; in-work <sup>a</sup> ; anchored <sup>a</sup> ; minimum income threshold <sup>c</sup> ; absolute <sup>c</sup> ). Jobless households <sup>a</sup> (x 2). Economic strain <sup>b</sup> . Deprivation (educational items <sup>b</sup> ; environment <sup>b</sup> ; household <sup>b</sup> ; ICT outside school <sup>c</sup> ; services <sup>c</sup> x 2; services <sup>b</sup> ; care services <sup>b</sup> )	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)
<b>Educational Well-being</b>	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 <sup>c</sup> ; literacy <sup>a</sup> ). Educational attainment (early leavers <sup>a</sup> ; failure rate in lower secondary <sup>c</sup> ; pass rates secondary <sup>c</sup> ; failure rate in secondary <sup>c</sup> ). Educational participation (truancy <sup>b</sup> )	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 3.2: An overview of indicators used in comparative child well-being analysis (cont.)**

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

97. 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.

98. In this comparison all indicators (regardless of how the authors originally categorised the data) are categorised into one of 10 dimensions: (i) health and safety, (ii) subjective well-being, (iii) personal family and peer relationships, (iv) material well-being and economic security, (v) educational well-being, (vi) behaviours and risks, (vii) housing and the local environment, (viii) quality of school life, (ix) family forms and care and (x) social, economic and civic participation.<sup>15</sup> The most commonly used indicators cover health (93 in total), behaviours and risk (61), and material well-being (47) dimension, almost two-thirds of all indicators in total. After these three core clusters, educational measures are then most commonly used (36), followed by housing and environmental indicators (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.

99. 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/records for young children (low birth-weights, immunisations, etc.) and educational enrolment amongst other aggregate statistics. The remaining 208 indicators were derived from surveys including these 8 studies that this review focuses on: 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).

100. Table 3.3 outlines the most commonly used indicators in the projects that are derived from survey data only. In the vast majority of cases (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 status, or social gradient (OECD, 2009, provides breakdown in socio-demographic status to some extent).

101. 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 for 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.

102. 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 (except ICCS, as mentioned above), and each of the dimensions of child well-being, will be covered in the analysis.

---

<sup>15</sup> 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.

**Table 3.3: 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 different child well-being frameworks.

### 3.3 Analysis of bias and examples from the literature

#### 3.3.1 Analysis of system-based bias

103. 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 were analysed for their effect on key child outcomes.

104. 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 (thus 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 rates 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 experience, 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 (Hanuschek and Woessmann, 2010).

105. Hanuschek and Woessmann (2010) also suggests 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 (for example, see discussion regarding exclusion of intellectually disabled children in sample design in section 2.1.3.3).

##### 3.3.1.1 Indicators and relationships with dropout rates and enrolment rates

106. In order to assess the extent to which enrolment rates in schools and dropout rates (by sex) between cohorts might affect the final results reported in any given survey, survey data across waves are pooled in time series with information on the target population (enrolment and dropout rates) derived from

education accounts. Tables 3.4 and 3.5 report the correlation analysis (and sensitised results) and show, for PISA, rates of enrolment are associated to average reading literacy scores as well as deprivation scores over the waves.

107. Three PISA variables are included in the analysis. The first is reading literacy achievement. PISA also collects literacy scores for science and mathematics, but 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 90<sup>th</sup> percentile result over the score at 10<sup>th</sup> percentile (used in both OECD, 2009 and UNICEF 2010), and the range of the mean standard error in 2000, 2003, 2006 and 2009.

108. The second PISA child well-being indicator concerns educational deprivation scores (used in 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.

109. 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 the European Index (Bradshaw et al., 2007). The measure has since been discontinued by PISA, after it was excluded from the compulsory questionnaire in 2003<sup>16</sup> and reinstated in 2006.

110. 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. 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 13 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 tests net of migration and immigration.

111. Table 3.4 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 thus in line with the findings of Hanuschek and Woessmann, 2010) and lower levels of educational deprivation and future work aspirations. Weaker associations are seen for distributional measures of 90<sup>th</sup>/10<sup>th</sup> percentiles in reading literacy, and the standard deviation for job aspirations, although directions of association follow those of similar indicators above.

---

<sup>16</sup> Countries that did not complete the question in the optional questionnaire in 2003 included: Canada, Denmark, Finland, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland and Turkey.

**Table 3.4: 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

	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 <sup>**</sup>	.518 <sup>**</sup>	.549 <sup>**</sup>	.534 <sup>**</sup>	.285 <sup>*</sup>	0.161	0.169	0.172
90/10 percentile ratio	-.202 <sup>*</sup>	-.220 <sup>*</sup>	-.210 <sup>*</sup>	-.200 <sup>*</sup>	-.316 <sup>**</sup>	-0.149	-0.080	-0.123
Reading SE range	-.196 <sup>*</sup>	-0.143	-0.129	-0.139	-0.190	-.204 <sup>*</sup>	-.205 <sup>*</sup>	-.228 <sup>*</sup>
Educational deprivation	-.731 <sup>**</sup>	-.512 <sup>**</sup>	-.519 <sup>**</sup>	-.525 <sup>**</sup>	-0.159	-0.050	0.060	-0.008
Aspired employment at 30 (ISEI scale)	-.583 <sup>**</sup>	-.486 <sup>**</sup>	-.475 <sup>**</sup>	-.497 <sup>**</sup>	-.289 <sup>*</sup>	-0.128	-0.150	-0.147
Aspired employment at 30 (ISEI scale - STDEV)	.508 <sup>**</sup>	.337 <sup>**</sup>	.266 <sup>*</sup>	.328 <sup>**</sup>	.399 <sup>**</sup>	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 inverted.

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

112. Because the range of gross enrolments of 15 year olds within Europe and the OECD vary to a large extent (between 102.2<sup>17</sup> 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).

113. The four right-hand columns of Table 3.4 show that many of the previously significant results have weakened (as for most enrolment rate estimates) or disappeared entirely (for dropout rates estimates) for all but one of the outcome measures. There are now no significant reports for deprivation scores.

114. After sensitivity test associations only three correlations strengthened. The size of the error range in reading literacy is now marginally significantly associated to dropout rates, suggesting that high enrolment countries with fewer dropouts also have a lower variation in reading literacy results. A result that is reflected by an increase in the strength of the association between the percentile range ratio and enrolment rates after sensitivity testing.

115. Table 3.5 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 (as these two countries can be considered outliers due to low scores and low

<sup>17</sup> Gross enrolment figures can be above 100% when enrolment exceeds population due to net migration and repeated years of study by pupils.

enrolment rates), 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 3.5: 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 enrolment 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**	0.272	0.194	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)	...	...	...	...	...	...	...	...

116. On the basis of these results, it is recommended that when reporting child well-being indicators using aggregate statistics from school data, tests for enrolment biases should be performed on a case-by-case basis and results for countries with significantly lower than average enrolment rates may be flagged for cautious interpretation if necessary.

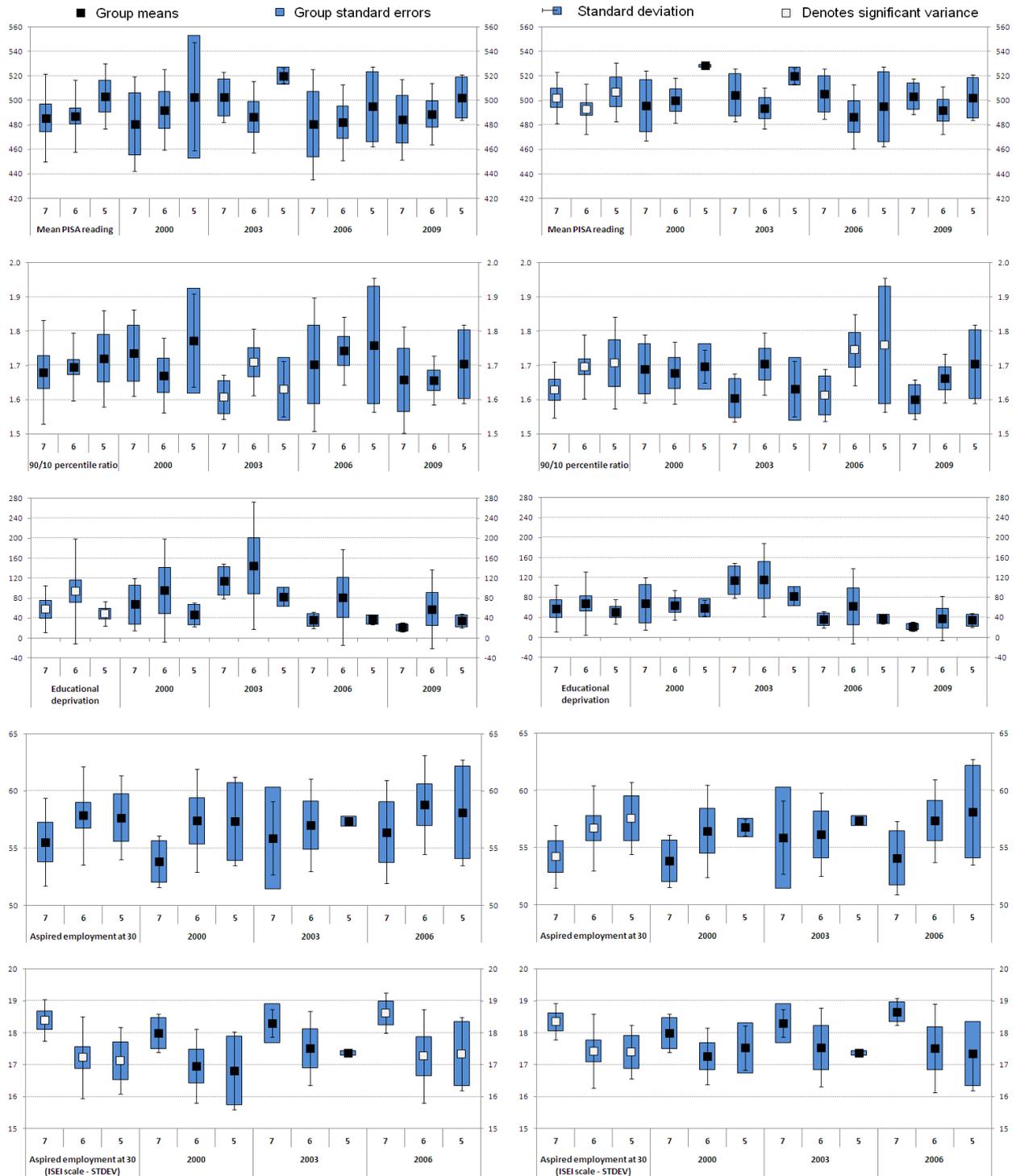
### 3.4.1.2 Indicators and relationships with school starting age

117. Across the OECD and Europe, compulsory schooling starts 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, policies which have 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 practices, including examinations, the school environment and their peers, and as such may better represent capabilities and experiences accumulated by the child than age itself.

118. Figure 3.1 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 the between-group variance (that is, the distributions of groups of countries starting school at different ages are significantly different from each other).

**Figure 3.1: 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 for countries grouped by school starting age (left-hand charts represent all countries; right-hand charts represent countries with enrolment rates above 90%)



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

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

119. 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 that 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 the square symbol for mean reading literacy difference in 2000 remains black there is no evidence (from analysis of variance tests – ANOVA) that the distribution of reading literacy in countries with different starting school ages are significantly different.

120. Results for the reading literacy indicators (mean scores and 90th/10th percentile) do not show consistent patterns for years-of-school related gains across the waves. Small, significantly different, results are only found between starting ages 6 and 7 years in literacy distribution scores in 2003 (before sensitivity) and 2006 and overall (after sensitivity) and significant differences between starting ages 5 and 6 years in mean 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.

121. 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 age 6 years have higher deprivation, but the difference is not significant. These results are quite clearly driven by low enrolment countries, as deprivation results are much more closely aligned in the sensitised tests.

122. 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 aspirations 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.

### ***3.3.2 Sampling and methodological bias***

123. 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 greatly across OECD countries and so sample sizes which rely on absolute numbers of children can affect comparability, statistical reliance on weighting methods, and techniques applied to infer confidence in population estimates. How and when questions are asked to children is likely to affect who answers these questions, affecting the accuracy and comparability of estimates country to country, and survey wave by survey wave.

124. 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 –for example 70%, 80% or 90% 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.

125. 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 (Micklewright et al., 2010). From this evidence the authors argue that there are serious limitations in using an arbitrary response rate threshold as a guide to quality.

126. 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 Micklewright et al. (2010) point out, 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.

127. The analysis below tests for significant associations between sample sizes and participation rates in school surveys and key well-being indicators, it tests for evidence of response attrition and bias in response attrition in background questionnaires, and tests for non-response by key child well-being indicators.

#### *3.3.2.1 Associations between sampling rates / populations / participation rates and child well-being indicators*

128. 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.

129. Table 3.6 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 population estimates. Following exclusions for the sensitivity tests (exclusion of lower enrolment countries as before) higher participation is weakly associated with lower deprivation, and higher participation is associated to the range of aspired work outcomes.

**Table 3.6: 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 enrolment 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 <sup>-</sup>	.247 <sup>-</sup>	-.222 <sup>*</sup>	0.167	-0.121	-.230 <sup>*</sup>
Standard error range	.237 <sup>-</sup>	-.287 <sup>-</sup>	-.452 <sup>-</sup>	.334 <sup>-</sup>	-.327 <sup>-</sup>	-.428 <sup>-</sup>
Aspired employment at 30 (ISEI scale)	.380 <sup>-</sup>	.303 <sup>-</sup>	-0.038	.283 <sup>*</sup>	0.188	0.040
Aspired employment at 30 (ISEI scale - STDEV)	-.308 <sup>-</sup>	-0.048	.330 <sup>-</sup>	-0.175	-0.005	.378 <sup>-</sup>

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

130. Table 3.7 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.

**Table 3.7: 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 enrolment 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 Secretariat calculations, OECD PISA (various waves), and OECD Population database (2011).

131. 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 reflection.

### 3.3.2.2 Organisation of items in the questionnaires and the effects of attrition

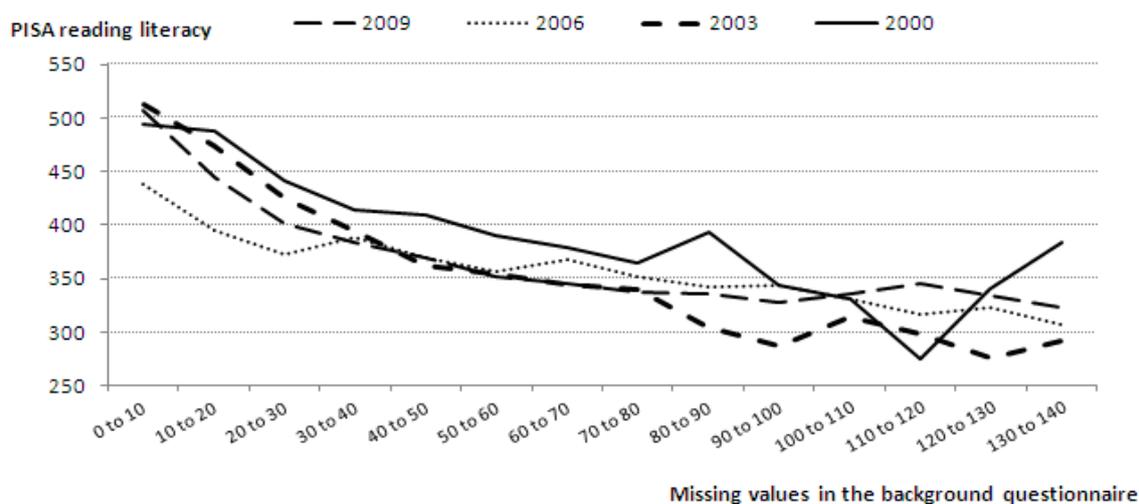
132. In educational assessment surveys, data derived from background questionnaires are essential for interpreting the variation in achievement indicators, and for developing additional indicators such as cultural possessions, home wealth, and family form. How a questionnaire is organised is likely to 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).

133. Figure 3.2 presents the combined OECD and EU country data for the four PISA waves. Trends are presented instead of correlations (as in Tables 3.4 – 3.7) 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 represents a concern for researchers and policy makers using information from background questionnaires in educational surveys, either to model variation or to indicate general (student) population estimates on any given background item. There is a concern for survey coordinators also, particularly in terms of where to place the most important explanatory factors when designing questionnaires.

**Figure 3.2: 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 Secretariat calculations, OECD PISA (various waves), and OECD Population database (2011).

134. Tests and adjustments for this type of response attrition are necessary by country (see Figure A5.1 in Annex 5 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. 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.

135. The data shown in Figure 3.2 does not map where in the questionnaire the missing values are, which is also likely to affect final response rates. To assess this effect, two tests were undertaken. The first test looks at missing values for some key child well-being measures to see whether some items are more sensitive to non-response than others (see Section 3.3.2.3). The second test assesses whether the position of an item in a questionnaire imposes non-response bias using aspirations data and parental qualifications data (see Section 3.3.2.4). In both cases these data are collected using open-ended questions which have not changed over the waves.

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

136. Table 3.8 compares literacy scores between respondents and non-respondents for nine indicators from the PISA 2009 survey (derived from different dimensions and domains as outlined above) for missing values and particular biases in terms of ability of respondents. The two purposes of this analysis are to (i) identify bias in the aggregated responses of the indicators themselves (for issues in this case of cognitive

ability to respond) and (ii) assess who might be implicitly excluded from bivariate or multivariate analysis using these data as dependent variable in assessments of outcomes.

137. Children with missing values always perform worse on average on the PISA reading literacy tests scores. 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’ have on average around a 100 point deficiency in the PISA scale (or broadly, one standard deviation) compared with pupils with responses. The smallest differences are seen in the fathers and mothers education levels, and whether ‘father lives at home’; indicators which also have the highest rates of missing values (see Annex 6 for disaggregated results by sex and migrant status).

**Table 3.8: Significant differences in means and variances between missing and non-missing values in selected PISA 2009 indicators**

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

138. It is clear from this evidence that children who are most likely to provide a useable response to the items do better in the PISA literacy test than children who fail to respond and are thus often excluded from contextual analysis. Moreover, and in the case of indicators commonly used in advanced analysis of the datasets, this also leads to exclusion from analysis designed to identify the effect sizes (for example, regression or fixed effect analysis).

139. To overcome this form of bias it is important to include variables that represent missing values in the analyses. For instance, parental education variables in the United Kingdom should include a nominal binary matrix designed to capture variations from a reference category (for example, parents having higher levels of education) which accounts for the missing category, and ensure that a minimum proportion of students from each category are included. However, it is unlikely that one fixed arbitrary cut-off point for minimum student participation in a category would be most appropriate for the treatment of all countries (such as under 10%), and rather that the analysis should include if children in the missing category have significantly different means and variances on the given dependent variable. This is also the case even if the sample size for this category is small and requires further grouping at a higher level. To this end the inclusion of the effect of the instrument should be preferred over the implicit bias introduced by excluding a set of cases with measurable non-response bias.

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

140. 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?

141. Figure 3.3 presents the differences in weighted and unweighted missing values for the work aspiration item between the PISA 2000 wave and the 2006 wave.<sup>18</sup> Although non-response is generally high (probably as the item 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.

142. 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 the outcome, the reasons behind this high level of non-response, and potential biases therein, need further investigation (it is possible that Ultraorthodox Jewish children do not have high job aspirations due to their religious commitments).

143. 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.

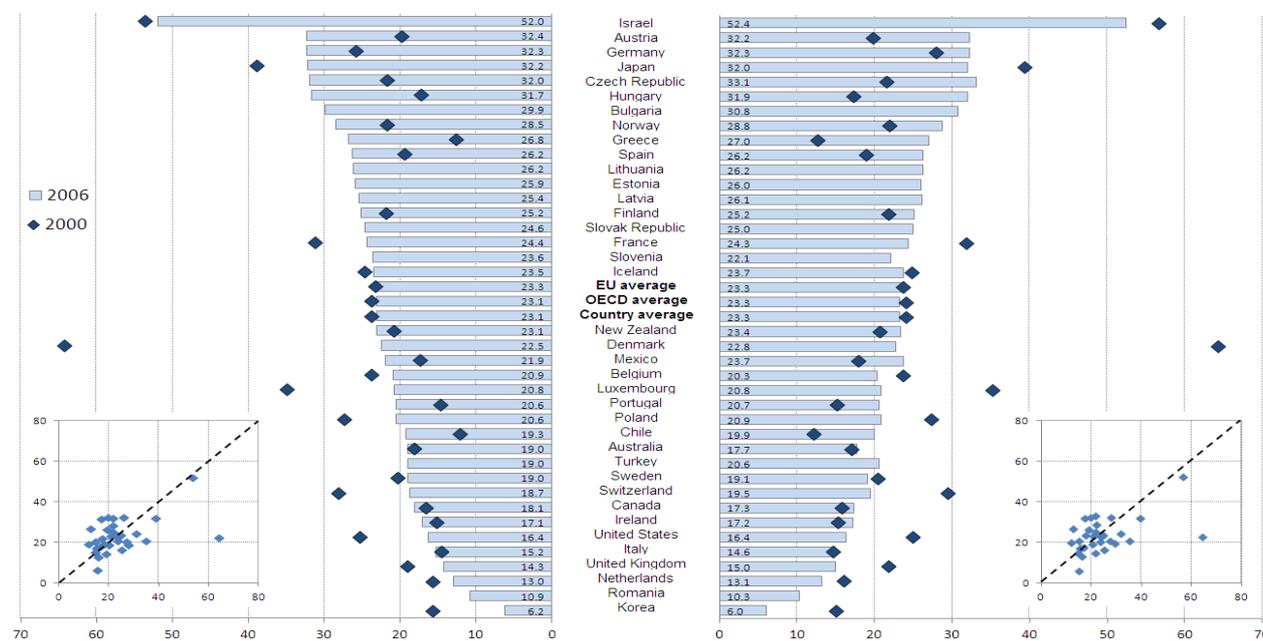
---

<sup>18</sup>

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.

**Figure 3.3: There is no clear pattern of changing non-response rates in aspirations between the waves**

Weighted (left-side) and unweighted (right-side) non-response rates for job aspirations questions in the 2000 and 2006 PISA background questionnaires



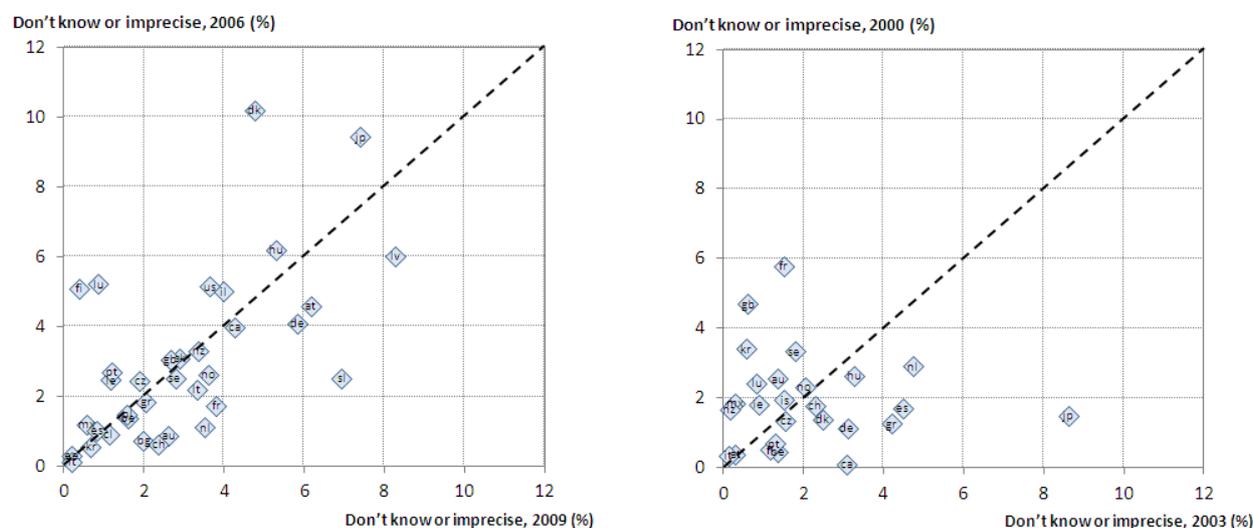
Source: OECD Secretariat calculations, OECD PISA (various waves).

144. Changes in the ordering of questions might not only affect the response-rates of indicators but may also affect the accuracy of responses, particularly for more difficult questions. 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 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.

145. 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). 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.

146. Figure 3.4 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 are the results from the 2000 and 2003 tests, showing that identically placed questions can produce big changes in the rate of imprecise answers between the two waves.

**Figure 3.4: Evidence suggests that the placement of complex items in background questions does not systematically affect the precision of responses on parental employment**



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

### 3.3.2.5 Indicators non-response bias in PISA

147. 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 A5.1 – A5.4.

148. What is evident 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 Section 5.5.2 for an example and caveats). These adjustment weights will have to be applied on a country by country basis, and on an indicator-by-indicator basis depending on groupings on non-response, in different countries across the different indicators.

149. In 2000, weighted data for parental interaction produced only 113 examples of countries with significant non-response by sex, 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.

### 3.3.2.6 Indicators non-response bias in HBSC

150. 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.<sup>19</sup> 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.

151. 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 less than one percent of students abstained from questions on bullying and enjoyment of school.

152. 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 presented below (for detailed results for all countries, see Annex 9).

#### 3.3.2.6.1 Frequency and correlates of non-response to the “life satisfaction” query

153. What factors make some students more likely than others to answer questions about their life satisfaction? Four potential correlates of non-response are evaluated here: sex, age, wealth, and health.<sup>20</sup> These four independent variables are selected because they are key attributes used in interpreting the findings of the study in international reports, and evaluated in a binary logistic model in which the outcome is a dummy variable (‘0’ or ‘1’) for the likelihood of response to the life satisfaction question (‘1’), and are presented in Table A9.1 of Annex 9. Based on a 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 were tabulated.

154. 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. The 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.

155. Looking across all of the 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.

<sup>19</sup> 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.

<sup>20</sup> 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 you say your health is excellent / good / fair / poor?”. These measures are reverse-coded with higher numbers representing poorer outcomes.

156. 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).

157. In contrast, 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

158. 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 (see next section). Finally, obesity in childhood has increasingly come to the attention of policymakers in a number of countries (such as the United Kingdom and the United States), and so reliable and representative measures of BMI are needed to analyse and inform child policy development across OECD countries.

159. 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.

160. 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.

161. 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.

162. 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 Germany and the United States, where self-reported “less healthy” students are more likely to answer the weight question.

163. 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 four of the countries: Belgium, Germany, Latvia, and the United States. High-income students are more likely to answer the weight question in the other countries: the Czech Republic, Greece, Lithuania, and the Netherlands.

164. The height question is presented adjacent to the weight question and is asked early-on in the survey; weight is the 13th question and height is the 14th question (out of over 100) in the HBSC surveys. The same four likely correlates of non-response are again evaluated: 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.

165. 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.

166. 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.

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

### 3.3.3 Cultural and linguistic bias

168. Blum et al., (2001), in an assessment of the International Adult Literacy Survey, found 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’ (Blum et al., 2001) is unsafe. The authors uncovered 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.

169. In comparing French with English translations a lack of precision in translation of questions leads to varied detail of responses. Blum et al., 2001, 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. 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. Evidence of this type of error was found for different English and French translations in IALS (Blum et al., 2001).

170. Errors in the IALS survey due to insensitivity to differences in culture knowledge were also evident. For example, cultural bias could also enter the results for a question which includes 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.

#### 3.3.3.1 Linguistic bias in translations in PISA

171. Linguistic tests are difficult to undertake on the many different language questionnaires in PISA simultaneously and is a barrier to such tests in this type of project. However, 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).

172. 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 a country to modify the term and two waves did not this provides an opportunity to explore the effect of introducing colloquial terms in translation practices 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’).<sup>21</sup>

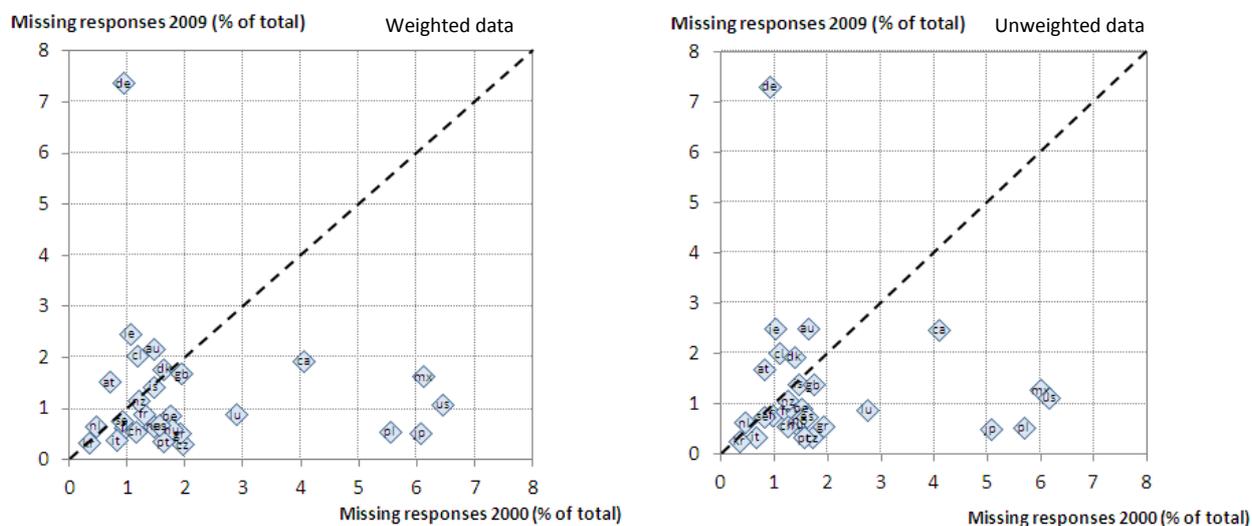
173. Figure 3.5 below plots 2000 against 2009 weighted and weighted non-response rate for ownership of cellular phones. Although non-response rates are small overall (unsurprising given the simplicity of the question), with more countries falling below the 45 degree line there is broad evidence to suggest 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 Japan, Mexico and the United States, response rates increased substantially. Only in Germany in 2009 have non-response rates increased to any degree.

174. The use of colloquial translations does not seem to have a systematic effect on the children’s non-response rates for 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 estimates should be excluded from comparisons and trends analysis if necessary.

---

<sup>21</sup> 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 in 2006 and 2009.

**Figure 3.5: In 2009 fewer German children provide useable responses to the cellular phone ownership question**



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

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

175. 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 a positive or negative incentive for disclosing birth place or heritage.

176. Table 3.9 shows the results for weighted non-response to the different migrant status questions for each member of the family from 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 3.9: 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	
<b>Average</b>	<b>2.02</b>	<b>1.65</b>	<b>-.37</b>		<b>.76</b>	<b>.99</b>	<b>.23</b>		<b>.73</b>	<b>1.14</b>	<b>.41</b>	

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

177. Comparing results shows 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 Japan, Israel and the United States, and higher non-response in Germany. For parents' migrant status, only five countries for mothers and three countries for fathers show improved response rates when country-specific options are available, although in no cases are shifts sufficiently large to suggest any form of significant bias in the results for parents' migrant status.

### 3.4 Comparisons of interchangeable measures: external and internal reliability

178. This section reviews key child well-being indicators that are repeated across the surveys as a final validity and reliability check on the indicators used in international comparisons of child well-being. Four comparisons are undertaken. The first tests the validity of the Family Affluence Scale produce by HBSC,

by recreating the composite measure using PISA data. The second assesses 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 compares data used for calculating poverty figures derived from household surveys (EU SILC and ESS), and the final section compares indicators of risk taking from the HBSC and ESPAD.

### 3.4.1 Family affluence in PISA and HBSC

179. Using the PISA 2000 survey, Family Affluence Scale (FAS) composites 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 in broad child well-being monitoring because 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.

180. 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.<sup>22</sup> 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).

181. HBSC Family affluence scale is computed as follows:

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

182. 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).

183. Table 3.10 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.

<sup>22</sup>

The response 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 the last two responses). The highest response across the four activities was recorded for each child.

184. The cross-plot in Figure 3.6 shows correlation coefficient of  $r=0.91$  ( $R^2$  of 0.82) between the PISA scale and the HBSC FAS scales findings<sup>23</sup>. The results here provide evidence of external validity in both the PISA family wealth items and the HBSC family wealth variable when composed in the form of a “summed scale”. This gives researchers 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).

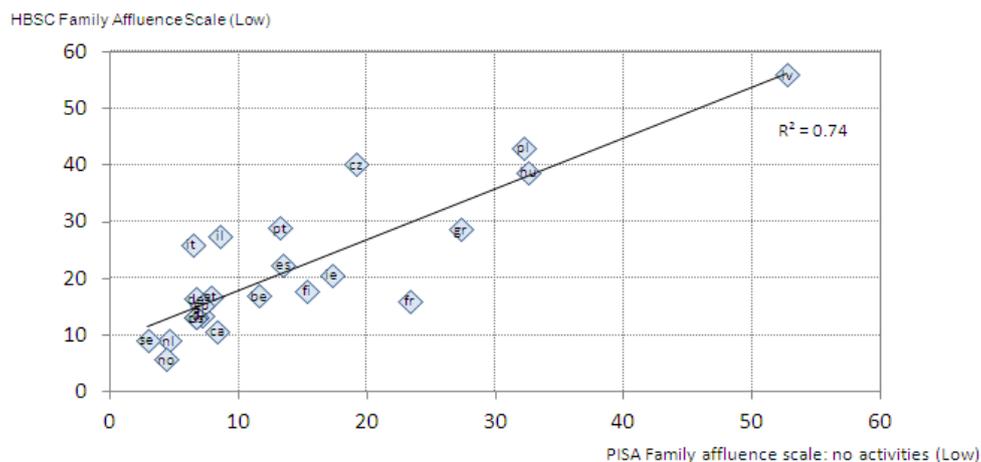
**Table 3.10: Percentage of youth population in each survey reporting low affluence by type of scale**

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
<b>Country average</b>	<b>8.8</b>	<b>14.6</b>	<b>22.0</b>

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

**Figure 3.6: Correlation between PISA FAS calculations and HBSC FAS calculations**



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

<sup>23</sup>

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  $r=0.88$ .

### 3.4.2 Literacy distributions in PIRLS and TIMSS

185. This section introduces comparisons of achievement indicator 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.

186. 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 A7.1 for TIMSS and Figure A7.2 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.

187. 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 distributions for Chile and Turkey are positively skewed. For the four countries with significant skews in the distribution 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 which has persisted across the survey waves.

188. 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.

189. In contrast to TIMSS, data for PIRLS is 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 8<sup>th</sup> grade pupils, while the PIRLS survey measures the reading ability across 4<sup>th</sup> grade pupils).

190. 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 (Figure A7.2 in Annex 7). The underlying data behind the chart does indeed confirm that the standard deviation for 2001 is a slightly less than 10.

191. 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 researchers 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.

192. 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 missing and/or hidden populations, see Section 4.2.1) 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.

### 3.4.3 Self-reported household structure and income in EU SILC and ESS

193. This section compared self-reported household structure and income in ESS and EU-SILC surveys. However, basic differences between the methodology of the ESS and EU-SILC 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 round 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<sup>24</sup>, Lithuania and the Slovak Republic) meant 21 countries are included in the analysis.<sup>25</sup>

#### 3.4.3.1 Comparing family structures

194. 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.

195. Figure 3.7 plots the proportion of three 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 very similar (see bottom right-hand chart of Figure 7). 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<sup>26</sup>, 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 (the line which represents a perfect match), though notably the trend for sole parents is much more uniform suggesting greater reliability, if not validity, of estimates for this population. 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.

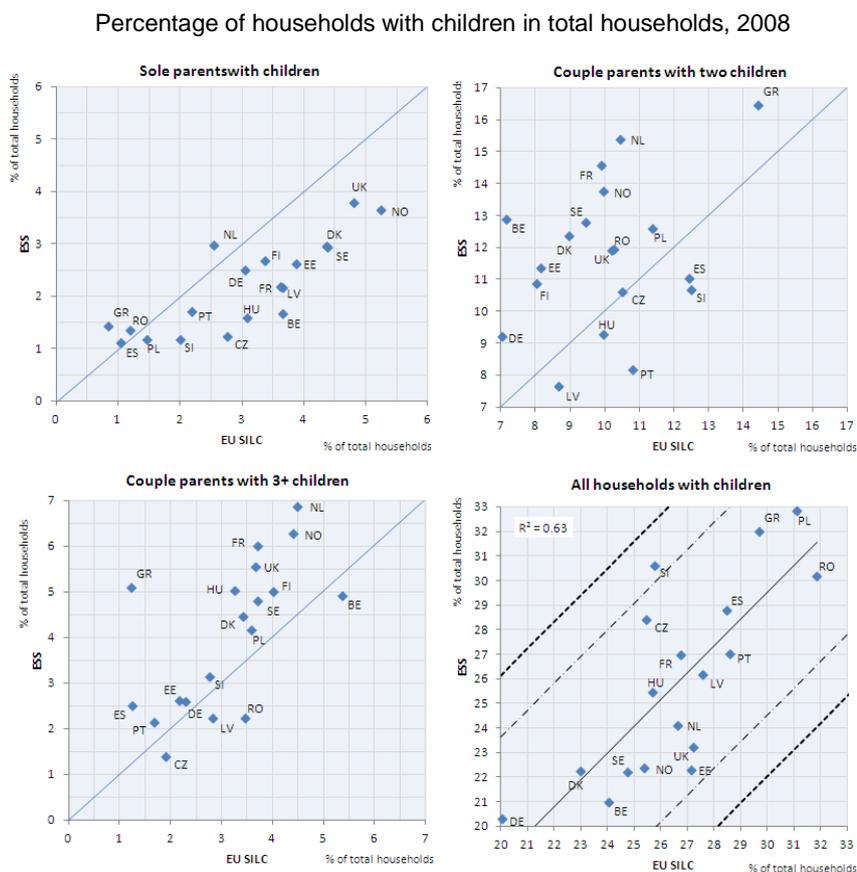
196. Although results are not encouraging, with the exception of couple families with 2 children, most differences could fall into margins of acceptable error on single indicators (plus 3% or minus 3%). Of concern, however, is that combined error by type will provide a very different picture of families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

<sup>24</sup> See footnotes 1 and 2 on page 10.

<sup>25</sup> 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.

<sup>26</sup> The overwhelming tendency would be to have greater confidence in EU-SILC figures due to the methodological strengths it has over ESS and larger sample sizes (see Section 2.1.3), which explains the order of comparison.

**Figure 3.7: Proportion of households with children is similar in ESS and EU SILC**



Source: OECD Secretariat calculations of ESS 2008 and EU-SILC 2008.

### 3.4.3.1 Comparing self-reported income in deciles

197. 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 the ranges reported in ESS, making the distributions comparable.

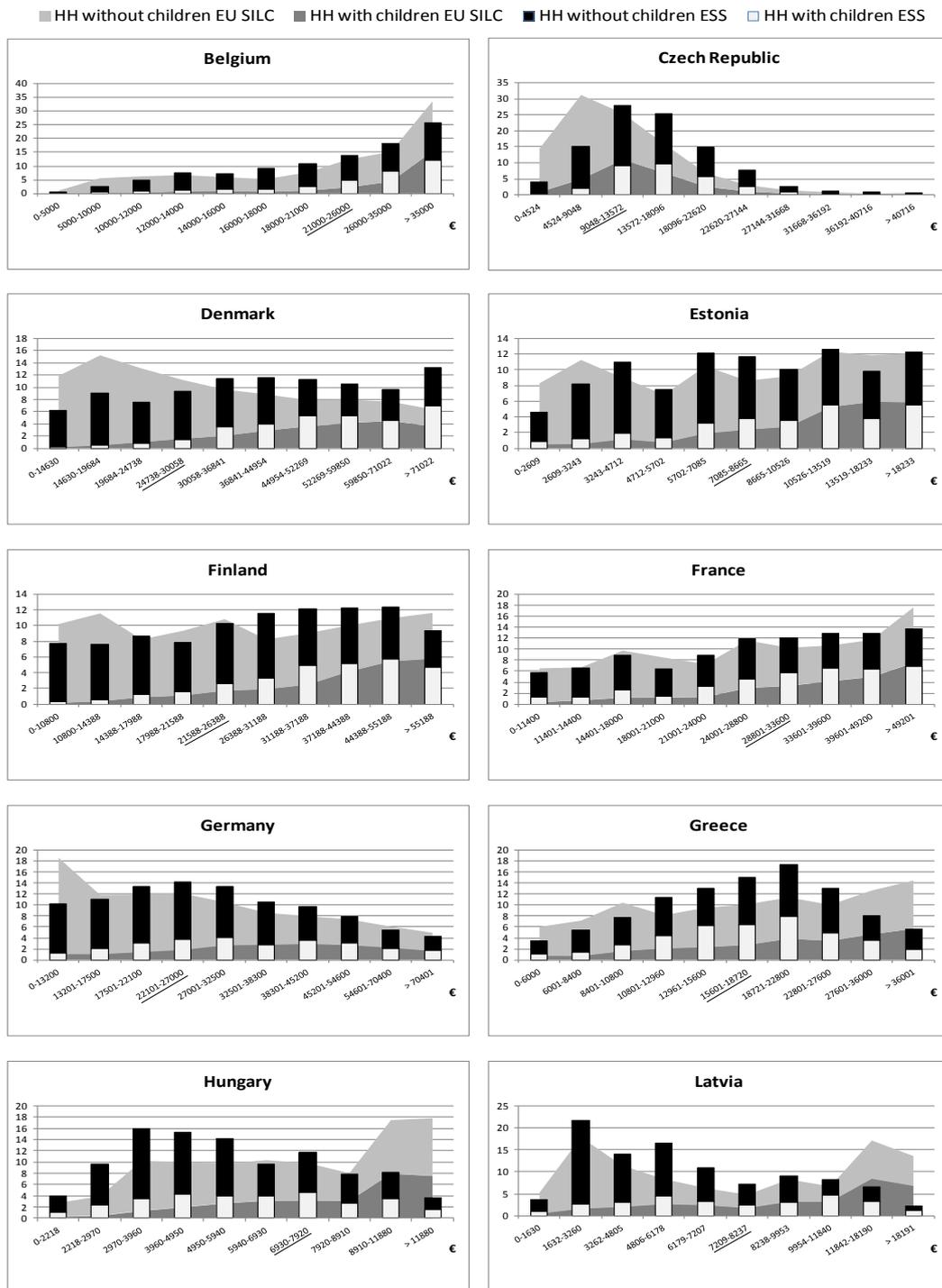
198. Figure 3.8 compares the distributions of household income in ESS and EU SILC for the 19 countries. On each chart, the proportions of ESS respondents in each income bracket are represented by the bar chart, with EU SILC proportions represented in the area behind. Proportions are broken down by households with children and households without children, the reason being that both groups' incomes contribute to the calculation of the poverty threshold.

199. 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 shifted forward one group in EU SILC.

200. For the remaining eleven countries (Denmark, Finland, Germany, Greece, Hungary, Latvia, Norway, Poland, Portugal, Spain and Sweden) different income group distributions are seen in the ESS and EU SILC results. Some countries present discrepancies more precisely for one or two income groups, 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 while the percentages for upper groups are similar). But for some other countries, the distribution is completely different from one survey to the other. For example, in the case for Portugal where ESS results show a 'bell curve' and EU SILC show a flatter distribution across the groups.

201. Again results are not encouraging, and the presentation of grouped income data as accurate population estimates without adjustments is not recommended. Neither is the use of grouped data for the calculation of income poverty estimates (using a mid-range method), however broad.

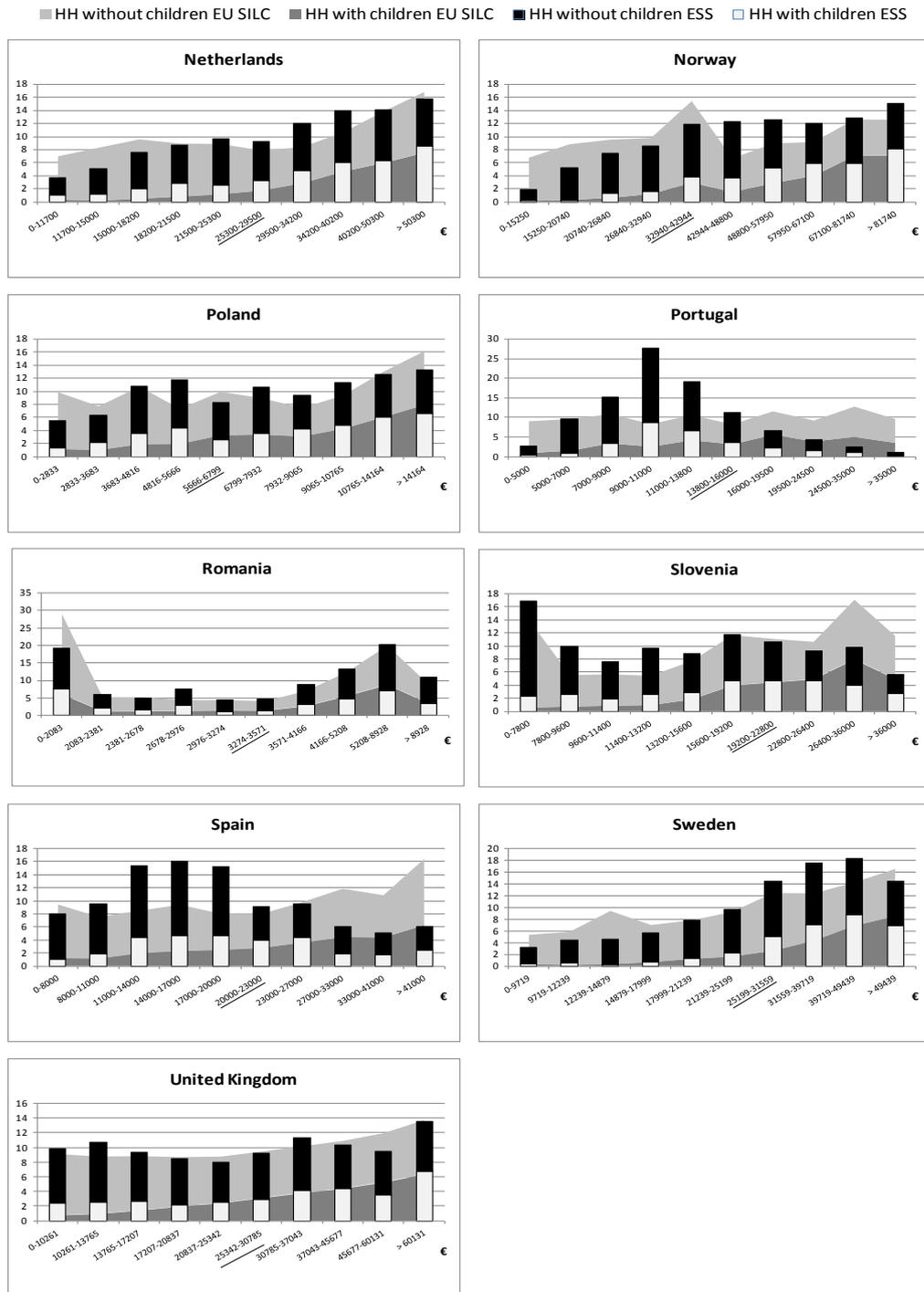
Figure 3.8: Household income distribution in ESS and EU SILC by household type, 2008.



Note: the median income is in the underlined group.

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

Figure 3.8. 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.

### 3.4.4 Smoking, drinking and drug use rates in HBSC and ESPAD

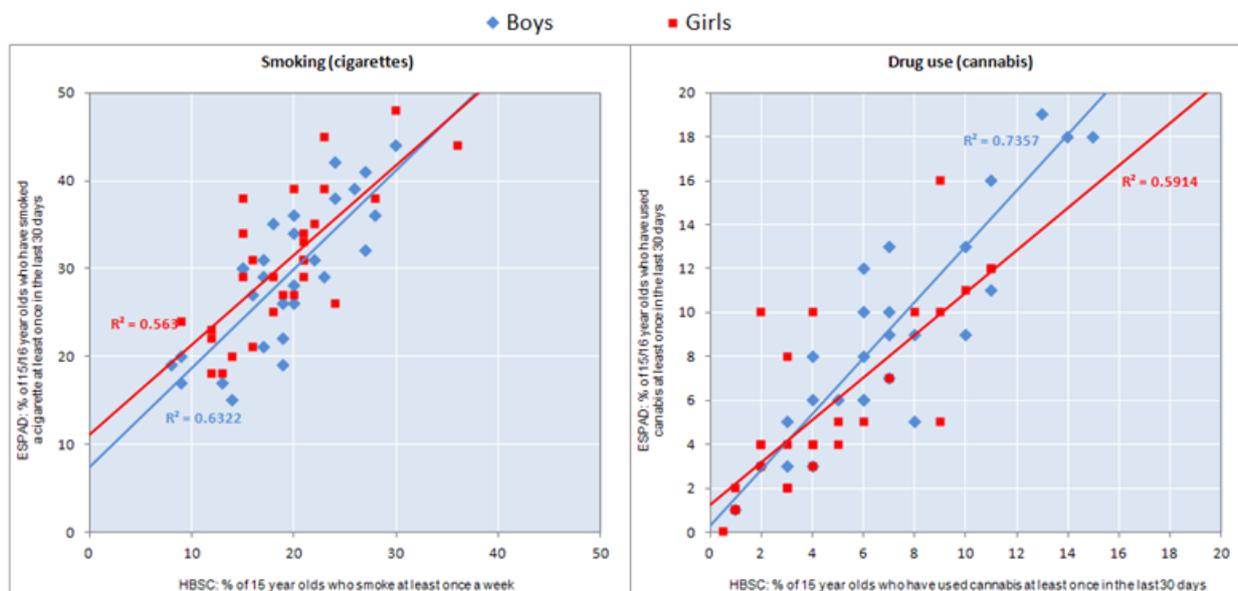
202. An important dimension of child well-being is risky behaviours, and especially substance abuse. Both the HBSC and ESPAD surveys cover a range of information on this subject, which can be used to compute comparable indicators on the prevalence of smoking cigarettes and cannabis use. These comparable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

203. 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 3.9. 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).

204. 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.

205. 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 tobacco use 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 3.9: Associations between smoking prevalence and cannabis use as reported in HBSC and ESPAD**



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

206. 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 than for cigarette or cannabis use, 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 difference in the definitions, the low level of correlation suggests these measures do not capture the same underlying risk factors.

### 3.5 Summary discussion

207. The empirical analysis of bias in child surveys has highlighted three key issues. Firstly, children's self-reported data are reliable between surveys and over time, and are robust to minor changes in method and question. Secondly, 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 taken into account when constructing indicators and subsequent analysis, including accounting for children that are not part of the survey (bias due to exclusion), and those children that are in the survey but are not responding to certain questions (bias related to non-response).

208. 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.

209. An important finding is that reliability in the responses 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 practice in data collection between child surveys). Changes to the same surveys across the waves also do not seem to dramatically change results.

210. Nevertheless, improvements can be made in the calculation of indicators for the purpose of international research and monitoring of child well-being (see Section 4.2). First, (whilst appreciating that many child well-being items are drawn from background questionnaires of surveys) the analysis shows that non-response affects the accuracy of estimates. Thus, comparisons and the adjustment to weights used to represent population estimates should be used to account for some of the bias resulting from non-response (see Section 4.2.2). Second, hidden populations, such as children not-enrolled in school, will affect the ability of the surveys to make unbiased population estimates. Much work can be done to better capture the hidden populations in child surveys (see Section 4.2.1).

## CHAPTER 4. RECOMMENDATIONS FOR THE USE AND IMPROVEMENT OF INTERNATIONAL DATASETS ON CHILDREN

211. 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 schools can be over burdened with surveys, it is important to make the best use of available international survey data on children.

212. This chapter is organised in two main parts. The first (section 4.1) 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 data for monitoring and policy informing. The second (section 4.2) outlines the key messages for the use and improvement of international datasets on children. The third (section 4.3) discusses which indicators are currently available for monitoring child well-being.

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

213. Using the examples of best practice discussed in Chapters 2 and 3 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, etc.) or other data used to identify the social gradient, such as the 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 child

well-being 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.

#### ***4.1.1 Recommendations for survey coordinators***

214. 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.

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

- Efforts should be made to improve communication between survey management groups and 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 used, 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 be the case for child well-being monitoring purposes in some cases). Moreover, good communication will enable researchers / monitors to plan future evaluations.
- Surveys should strive to achieve basic requirements for trends analysis, and consider ways to capture longitudinal experiences of schools, if not households or children themselves. With the exception of the PISA longitudinal survey (undertaken in 5 countries – Australia, Canada, the Czech Republic, Switzerland and Uruguay), 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 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 (their questionnaire has been unchanged over three waves – 2001/02, 2005/06 and 2009/10). Five surveys in total explicitly mention developing 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 4.1).
- 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. Advice, or guidelines, on how to develop weighting systems to account for non-response on key indicators would also be welcome (for example, see Section 4.2.2).
- All of the surveys are good at exploring their subject matter (education, health behaviours, and so on), but perhaps more limited at covering a full range of contextual factors to help explain their findings. This is likely to become increasingly important over time as the role of ‘softer’ measures in explaining what works in these area (such as parenting practices, family relationships and subjective well-being) gathers momentum. Almost all of the surveys mention informing policy as a key purpose of the work. Thus, 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 reflects 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 being 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 coverage by age in child surveys. Good practice guidelines, beyond household weights by number of children, are needed.
- Population weights which do not inflate sample numbers (N's) (that is, 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, 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 trends analysis) allows.
- More experimental methods could be used to get at validity issues in surveys, particularly around sensitive items (such as risk behaviours, or parental and peer relationships). A good example from ESPAD is the inclusion of 'relevis' in a list of drugs used by the respondent (sedatives, amphetamines, LSD, crack, cocaine, heroin, Psilocybin mushrooms, GHB and steroids are also included). Relevis is a dummy drug used to test children's honesty in answering these questions. Analysis of relevis responses can be used to adjust the findings of real, but sensitive, items included in the survey.

**Box 4.1: 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 et al. (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 and 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 be 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 its 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 et al. (2011) for more detail)).

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 4.2.3.

#### 4.1.2 Recommendations for the use of child well-being indicators

216. 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 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 (for example, see Section 4.2.2).
- For complex analyses, 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 (such as 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 efforts to develop inequality measures, expand dimensions coverage (from available data on 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 differently 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.

#### **Box 4.2: 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 a child's true education ability unless the child themselves responded to the tests. Moreover, when assessing health behaviours 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 feasible, 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 through 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 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 in longitudinal surveys of children's well-being, the child's parents or guardians will need to represent the children in such surveys.

In order to collect reliable and accurate child well-being data, that helps children achieve their rights, interviewing children is necessary, especially in circumstances where information collected attempts to reflect personal experiences, needs, abilities 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.

## **4.2 The way forward on three key issues**

### ***4.2.1 The Why and How of surveying hidden populations***

217. 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

excluded from the surveys (in some cases, countries are represented by areas, such as Germany and the United Kingdom being represented by smaller regions in the HBSC survey).

218. Using censuses and other collections methods to study at-risk populations can provide a picture of acute risks (e.g. low education levels and low income levels) 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 (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 have low school enrolment rates, questions are raised about their representativeness within indicators derived from international surveys, and importantly, ways in which surveys could capture these populations.

219. Although appropriate cost benefit analysis should be undertaken beforehand, several steps could be undertaken 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.

220. 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.

#### ***4.2.2 Using weights to adjust for non-response in indicator development***

221. 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 (see Section 3.3.2), adjustments for differences in non-response is essential for comparing indicators developed at a national level between countries and over time.

222. 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 elicit very different response rates (see tables A.5.1 – 5.4 in Annex 5 for examples across the four waves of PISA). Importantly for education surveys, achievement does associate with overall responses (Figure A5.1 in Annex 5). 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.

223. 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. Table 4.1 provides an example of adjustment for the underrepresentation of girls with low FAS scores on a given item, with an adjustment weight of 1.15 (inflating responses by 15% overall) 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 4.1: 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.

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

225. 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 and would throw up more factors than those which could realistically be adjusted by 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.

226. 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 in the PISA survey).

#### **4.2.3 Harmonising data collections**

227. The practicalities for harmonisation between child surveys have been outlined in Box 4.1. This section now goes on to suggest specific indicators which could be harmonised across surveys, and presents a rationale for their selection.

228. 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, the information should be part of 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 different surveys is produced in order to accurately match data.

##### *4.2.3.1 Harmonising socio-demographic measures*

229. Based on the above criteria, the first suggestion is to harmonise the surveys using data that is already used to collect socio-demographic information (as discussed previously in Section 2.2.1). Table 4.2 provides information on specific items that could be used to harmonise child surveys, and includes: age, sex, family form, and migrant status.

230. Each item is explained in detail, and taken either directly from one of the surveys or adapted from examples across surveys (see two most right-hand columns of Table 4.2).<sup>27</sup> 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.

---

<sup>27</sup> In both Tables 4.2 and 4.3 some simplified items, reflecting questions asked in more than one item across surveys, have been suggested. The reason for this is to meet the selection criteria outlined above, particularly in terms of applicability, simplicity and parsimony.

**Table 4.2: 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.

231. The recommendations here are naturally limited to what is considered to be feasible 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. Any final list would need to be agreed between survey providers and it does not, on its own, meet recommendation by the EC Social Protection Committee and others for an effective monitoring tool 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).

232. Of all of the recommendations in Table 4.2, 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 surveys children who are not presently living in a foster or children's home. Although this means sacrificing time and space in other surveys, it would seem to be very relevant considering the trend of growing single and reformed families, and as social policies designed to adapt for children living in more than one household are becoming more common.

#### 4.2.3.2 Harmonising socio-economic measures

233. Following efforts to come up with a minimal set of socio-demographic indicators to harmonise surveys, an effort could also be made to match socio-economic items of material deprivation and parental education and employment (see Table 4.3).

234. Child deprivation items are collected 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 added effort to surveys that lack such information (this is also less complicated than asking PISA to undertake questions on risk, or morbidity, and HBSC survey would be unlikely to ever undertake a literacy assessment type module). The added value of harmonising deprivation items would be the opportunity to expand substantially the analysis on social gradient (a key recommendation of this report and the SPC report (2012)).

235. Parental employment and education data are 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 4.3: Suggested socio-economic items for harmonisation across surveys**

Matching indicator	Subject or question	Item detail	Survey use similar item	Survey use exact item
<b>Deprivation</b>	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
<b>Parental education</b>	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)	
<b>Parental Employment</b>	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.

### 4.3 Which indicators can be used, now, for monitoring child well-being

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

237. As part of their on-going work on child poverty and child well-being, the Social Protection Committee (SPC) 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 and 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.

238. 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 and not 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 4.4) 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 regularly monitor the well-being of vulnerable children are to be met.

239. 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 2 (see Section 2.2.2) 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 4.3 and 4.4, 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 – 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.

240. In relation to **children's health and inequalities**, or health by social gradient, and acknowledging the many administrative series that could be used are not discussed in this survey report (the SPC report lists infant mortality, as well as health service indicators – although physical safety measures like intentional mortality should also be included), ESPAD and HBSC can be used to identify health indicators in the child population in Europe (from age 11 through to 16 years). 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 based on these sources. The indicators used in the UNICEF Innocenti Report Card 9 *The children left behind* (UNICEF, 2010) are listed in Table 4.4, 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 4.3 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 4.4: 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 costs	PISA	Did you attend pre-school <ISCED 0>	Yes/No:
		EU-SILC	Number of hours of education 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.

241. 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 (and ISCED 0) in their questionnaires. Using this information inference can be drawn about the quality of childcare by relating childcare attendance in PISA to literacy outcomes and, in relation to access, 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 time series derived from administrative data, or micro-simulation modelling.

242. **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 4.4 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, a precursor to ICCS, in 1999 collected information on whether children were actively participating in political and social groups (the ICCS data reported only covers opinions), and is an example for summary work looking to develop more participation items in future surveys/waves (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 on 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).

243. **Alternative poverty measures** mentioned in the SPC report (2012) refer to food and fuel poverty. Table 4.4 picks up on this suggestion and lists recent 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 thus risks to supplies of necessities for the home). The arrears and debt figures would complement efforts to measure deprivation and income poverty as they are not included in wealth or income measures (including poverty rates measures). 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.

244. 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, and income statistics, 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 time 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 4.1.

### ***5.6.1 Issues of coverage and timeliness in available monitoring data***

245. For each of the proposals above there are issues of country coverage and the timeliness of collection and reporting to consider. Country coverage varies by surveys, 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, then attempt to improve the coverage (concepts and countries) and timeliness of reporting during the process of monitoring. Otherwise, waiting for 'the perfect time' to begin

work may lead to important messages and lessons being missed that could inform policy and practice in the meantime.

246. Key steps to improve the coverage of child well-being indicators, and country coverage, are outlined in other parts of this paper, 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.

### 4.3 Summary

247. Despite good evidence in support of the robustness of much of the available child well-being data, established 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’ monitoring tool in evidence based policy for *all* children. Missing populations in the data (for example, children out of school or children under age 9 years), 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.

248. Although there remain gaps in what we know about how to improve the measurement of children’s well-being cross-nationally, this project has achieved a number of important steps. The review in Chapter 2 provides a systematic outline for recommending the adaptations to content in existing surveys, and also outlines the substantial gaps in knowledge in child well-being that remains within these surveys. Chapter 3 provides empirical grounding useful for understanding how amendments to the position or wording of items 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). This final chapter provides the first steps in efforts to understand and recommend methods for harmonisation in the field.

249. Whilst researchers and policymakers are waiting for new comparative data it is important that the best is made of data that are available now. 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 obvious and investment in new collections, and 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 implement some of the steps recommended in this final chapter, a future for effective monitoring of comparative child well-being, for the benefit of all children, can be secured.

## ANNEX

Please find the complete annex to the working paper at [www.oecd.org/social/family/database/CWBM](http://www.oecd.org/social/family/database/CWBM)  
(direct link: [www.oecd.org/els/family/Evaluation%20of%20Intl%20Surveys%20of%20Children%20-%20SEM%20No.146%20-%20Annex.pdf](http://www.oecd.org/els/family/Evaluation%20of%20Intl%20Surveys%20of%20Children%20-%20SEM%20No.146%20-%20Annex.pdf)).

## REFERENCES

- Blum, A., Goldstein, H., and F. Guérin-Pace (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 D. Richardson (2009), “An Index of Child Well-being in Europe”, *Child Indicators Research*, Vol. 2, No. 3, pp 319 - 351.
- Bradshaw, J., Hoelscher, P., and D. Richardson (2007). An Index of Child Well-being in the European Union. *Social Indicators Research*, 80(1), 133-77.
- Currie, C., Zanotti, C., Morgan, A., Currie, D., de Looze, M., Roberts, C., Samdal, O., Smith, O., and V. Barnekow (2008), “Inequalities in 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. Copenhagen.
- Currie, C., Roberts, C., Morgan, A., Smith, R., Settertobulte, W., Samdal, O. and V. Barnekow Rasmussen, (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/](http://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](http://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](http://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/](http://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](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 D. Richardson (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

Hanushek, 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 C. Skinner (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, Martin, M. O., Kennedy, A. M., Trong, K. L., and M. Sainsbury (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](http://www.oecd.org/els/social/childwellbeing)).

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](http://www.oecd.org/pages/0,3417,en_32252351_32235731_1_1_1_1_1,00.html)

OECD (2011) *Doing Better for Families*, OECD, Paris ([www.oecd.org/social/famil/doingbetter](http://www.oecd.org/social/famil/doingbetter)).

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](http://www.oecd.org/pages/0,3417,en_32252351_32236130_1_1_1_1_1,00.html).

OECD (2013). *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, Paris.

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

Rose, R., and Y. 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, 27<sup>th</sup> 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](http://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.childrensociety.org.uk/sites/default/files/tcs/good\\_childhood\\_report\\_2012\\_final.pdf](http://www.childrensociety.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.unhcr.ch/html/menu3/b/k2crc.htm](http://www.unhcr.ch/html/menu3/b/k2crc.htm).

UNESCO Institute for Statistics (2011): 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.

**OECD SOCIAL, EMPLOYMENT AND MIGRATION WORKING PAPERS**

Most recent releases are:

- No. 159 *TRENDS IN TOP INCOMES AND THEIR TAXATION IN OECD COUNTRIES*, Michael Förster, Ana Llana-Nozal and Vahé Nafilyan (2014)
- No. 158 *SKILLS AT WORK: HOW SKILLS AND THEIR USE MATTER IN THE LABOUR MARKET*, Glenda Quintini (2014)
- No. 157 *CHANGES IN FAMILY POLICIES AND OUTCOMES: IS THERE CONVERGENCE?*, Willem Adema, Nabil Ali, and Oliver Thévenon (*Forthcoming*)
- No. 156 *RETOUR À L'EMPLOI DES CHOMEURS SENIORS FRANÇAIS AYANT BÉNÉFICIE D'UN ACCOMPAGNEMENT RENFORCÉ VERS L'EMPLOI EN 2009 ET 2010*, Gwenn Parent (2014)
- No. 155 *MIGRATION AS AN ADJUSTMENT MECHANISM IN THE CRISIS? A COMPARISON OF EUROPE AND THE UNITED STATES*, Julia Jauer, Thomas Liebig, John P. Martin and Patrick Puhani (2014)
- No. 154 *SAME BUT DIFFERENT: SCHOOL-TO-WORK TRANSITIONS IN EMERGING AND ADVANCED ECONOMIES*, Glenda Quintini and Sébastien Martin (2014)
- No. 153 *A NEW MEASURE OF SKILLS MISMATCH*, Michele Pellizzari and Anne Fichen (2013)
- No. 152 *CATASTROPHIC JOB DESTRUCTION*, Anabela Carneiro, Pedro Portugal and José Varejão (2013)
- No. 151 *THE PERVERSE EFFECTS OF JOB-SECURITY PROVISIONS ON JOB SECURITY IN ITALY: RESULTS FROM A REGRESSION DISCONTINUITY DESIGN*, Alexander Hijzen, Leopoldo Mondauto, Stefano Scarpetta (2013)
- No. 150 *REDISTRIBUTION POLICY IN EUROPE AND THE UNITED STATES: IS THE GREAT RECESSION A 'GAME CHANGER' FOR WORKING-AGE FAMILIES?* Herwig Immervoll, Linda Richardson (2013)
- No. 149 *A BIRD'S EYE VIEW OF GENDER DIFFERENCES IN EDUCATION IN OECD COUNTRIES* Angelica Salvi Del Pero and Alexandra Bytchkova (2013)
- No. 148 *TRENDS IN POVERTY AND INEQUALITY IN DECENTRALISING INDONESIA* Riyana Miranti, Yogi Vidyattama, Erick Hansnata, Rebecca Cassells and Alan Duncan (2013)
- No. 147 *WOMEN ENTREPRENEURS IN THE OECD: KEY EVIDENCE AND POLICY CHALLENGES* Mario Piacentini (2013)
- No. 146 *AN EVALUATION OF INTERNATIONAL SURVEYS OF CHILDREN*, Dominic Richardson and Nabil Ali (2014)
- No. 145 *DRIVERS OF FEMALE LABOUR FORCE PARTICIPATION IN THE OECD* Olivier Thévenon (2013)
- No. 144 *THE ROLE OF SHORT-TIME WORKING SCHEMES DURING THE GLOBAL FINANCIAL CRISIS AND EARLY RECOVERY*, Alexander Hijzen, Sébastien Martin (2012)
- No. 143 *TRENDS IN JOB SKILL DEMANDS IN OECD COUNTRIES*, Michael J. Handel (2012)
- No. 142 *HELPING DISPLACED WORKERS BACK INTO JOBS AFTER A NATURAL DISASTER: RECENT EXPERIENCES IN OECD COUNTRIES*, Danielle Venn (2012)
- No. 141 *LABOUR MARKET EFFECTS OF PARENTAL LEAVE POLICIES IN OECD COUNTRIES*, Olivier Thévenon & Anne Solaz (2012)
- No. 140 *FATHERS' LEAVE, FATHERS' INVOLVEMENT AND CHILD DEVELOPMENT: ARE THEY RELATED? EVIDENCE FROM FOUR OECD COUNTRIES*, Maria C. Huerta, Willem Adema, Jennifer Baxter, Wen-Jui Han, Mette Lausten, RaeHyuck Lee and Jane Waldfogel (2012)
- No. 139 *FLEXICURITY AND THE ECONOMIC CRISIS 2008-9 – EVIDENCE FROM DENMARK*, Tor Eriksson (2012)

- No. 138 *EFFECTS OF REDUCING GENDER GAPS IN EDUCATION AND LABOUR FORCE PARTICIPATION ON ECONOMIC GROWTH IN THE OECD*, Olivier Thévenon, Nabil Ali, Willem Adema and Angelica Salvi del Pero (2012)
- No. 137 *THE RESPONSE OF GERMAN ESTABLISHMENTS TO THE 2008-2009 ECONOMIC CRISIS*, Lutz Bellman, Hans-Dieter Gerner, Richard Upward (2012)
- No. 136 Forthcoming *THE DYNAMICS OF SOCIAL ASSISTANCE RECEIPT IN GERMANY* Sebastian Königs
- No. 135 *MONEY OR KINDERGARTEN? DISTRIBUTIVE EFFECTS OF CASH VERSUS IN-KIND FAMILY TRANSFERS FOR YOUNG CHILDREN*, Michael Förster and Gerlinde Verbist (2012)
- No. 134 *THE ROLE OF INSTITUTIONS AND FIRM HETEROGENEITY FOR LABOUR MARKET ADJUSTMENTS: CROSS-COUNTRY FIRM-LEVEL EVIDENCE*, Peter N. Gal (VU University Amsterdam), Alexander Hijzen and Zoltan Wolf (2012)
- No. 133 *CAPITAL'S GRABBING HAND? A CROSS-COUNTRY/CROSS-INDUSTRY ANALYSIS OF THE DECLINE OF THE LABOUR SHARE*, Andrea Bassanini and Thomas Manfredi (2012)
- No. 132 *INCOME DISTRIBUTION AND POVERTY IN RUSSIA*, Irinia Denisova (2012)
- No. 131 *ELIGIBILITY CRITERIA FOR UNEMPLOYMENT BENEFITS*, Danielle Venn (2012)
- No. 130 *THE IMPACT OF PUBLICLY PROVIDED SERVICES ON THE DISTRIBUTION OF RESOURCES: REVIEW OF NEW RESULTS AND METHODS*, Gerlinde Verbist, Michael Förster and Maria Vaalavuo (2012)
- No. 129 *AN OVERVIEW OF AUSTRALIA'S SYSTEM OF INCOME AND EMPLOYMENT ASSISTANCE FOR THE UNEMPLOYED*, Peter Davidson, Peter Whiteford (2012)
- No. 128 *THE INTEGRATION OF IMMIGRANTS AND THEIR CHILDREN INTO THE LABOUR MARKET IN SWITZERLAND*, Thomas Liebig, Sebastian Kohls and Karoline Krause (2012)
- No. 127 *THE LABOUR MARKET INTEGRATION OF IMMIGRANTS AND THEIR CHILDREN IN AUSTRIA*, Karolin Krause and Thomas Liebig (2011)
- No. 126 *ARE RECENT IMMIGRANTS DIFFERENT? A NEW PROFILE OF IMMIGRANTS IN THE OECD BASED ON DIOC 2005/06*, Sarah Widmaier and Jean-Christophe Dumont (2011)
- No. 125 *EARNINGS VOLATILITY AND ITS CONSEQUENCES FOR HOUSEHOLDS*, Danielle Venn (2011)
- No. 124 *CRISIS, RECESSION AND THE WELFARE STATE*, Willem Adema, Pauline Fron and Maxime Ladaique (2011)
- No. 123 *AGGREGATE EARNINGS AND MACROECONOMIC SHOCKS* Andrea Bassanini (2011)
- No. 122 *REDISTRIBUTION POLICY AND INEQUALITY REDUCTION IN OECD COUNTRIES: WHAT HAS CHANGED IN TWO DECADES?* Herwig Immervoll, Linda Richardson (2011)
- No. 121 *OVER-QUALIFIED OR UNDER-SKILLED*, Glenda Quintini (2011)
- No. 120 *RIGHT FOR THE JOB*, Glenda Quintini (2011)
- No. 119 *THE LABOUR MARKET EFFECTS OF UNEMPLOYMENT COMPENSATION IN BRAZIL*, Alexander Hijzen (2011)
- No. 118 *EARLY MATERNAL EMPLOYMENT AND CHILD DEVELOPMENT IN FIVE OECD COUNTRIES*, Maria del Carmen Huerta, Willem Adema, Jennifer Baxter, Miles Corak, Mette Deding, Matthew C. Gray, Wen-Jui Han, Jane Waldfogel (2011)
- No. 117 *WHAT DRIVES INFLOWS INTO DISABILITY? EVIDENCE FROM THREE OECD COUNTRIES* Ana Llena-Nozal and Theodora Xenogiani (2011)
- No. 116 *COOKING, CARING AND VOLUNTEERING: UNPAID WORK AROUND THE WORLD*, Veerle Miranda (2011)

- No. 115 *THE ROLE OF SHORT-TIME WORK SCHEMES DURING THE 2008-09 RECESSION*, Alexander Hijzen and Danielle Venn (2010)
- No. 114 *INTERNATIONAL MIGRANTS IN DEVELOPED, EMERGING AND DEVELOPING COUNTRIES: AN EXTENDED PROFILE*, Jean-Christophe Dumont, Gilles Spielvogel and Sarah Widmaier (2010)
- No. 113 *ACTIVATION POLICIES IN JAPAN*, Nicola Duell, David Grubb, Shruti Singh and Peter Tergeist (2010)
- No. 112 *ACTIVATION POLICIES IN SWITZERLAND*, Nicola Duell and Peter Tergeist with contributions from Ursula Bazant and Sylvie Cimper (2010)
- No. 111 *ECONOMIC DETERMINANTS AND CONSEQUENCES OF CHILD MALTREATMENT* Lawrence M. Berger, Jane Waldfogel (forthcoming)
- No. 110 *DISTRIBUTIONAL CONSEQUENCES OF LABOR DEMAND ADJUSTMENTS TO A DOWNTURN: A MODEL-BASED APPROACH WITH APPLICATION TO GERMANY 2008-09*, Herwig Immervoll, Olivier Bargain, Andreas Peichl, Sebastian Siegloch (2010)
- No. 109 *DECOMPOSING NOTIONAL DEFINED-CONTRIBUTION PENSIONS: EXPERIENCE OF OECD COUNTRIES' REFORMS*, Edward Whitehouse (2010)
- No. 108 *EARNINGS OF MEN AND WOMEN WORKING IN THE PRIVATE SECTOR: ENRICHED DATA FOR PENSIONS AND TAX-BENEFIT MODELING*, Anna Cristina D'Addio and Herwig Immervoll (2010)
- No. 107 *INSTITUTIONAL DETERMINANTS OF WORKER FLOWS: A CROSS-COUNTRY/CROSS-INDUSTRY APPROACH*, Andrea Bassanini, Andrea Garnerò, Pascal Marianna, Sebastien Martin (2010)
- No. 106 *RISING YOUTH UNEMPLOYMENT DURING THE CRISIS: HOW TO PREVENT NEGATIVE LONG-TERM CONSEQUENCES ON A GENERATION?* Stefano Scarpetta, Anne Sonnet and Thomas Manfredi (2010)
- No. 105 *TRENDS IN PENSION ELIGIBILITY AGES AND LIVE EXPECTANCY, 1950-2050* Rafal Chomik and Edward Whitehouse (2010)
- No. 104 *ISRAELI CHILD POLICY AND OUTCOMES* John Gal, Mimi Ajzenstadt, Asher Ben-Arieh, Roni Holler and Nadine Zielinsky (2010)
- No. 103 *REFORMING POLICIES ON FOREIGN WORKERS IN ISRAEL* Adriana Kemp (2010)
- No. 102 *LABOUR MARKET AND SOCIO-ECONOMIC OUTCOMES OF THE ARAB-ISRAELI POPULATION* Jack Habib, Judith King, Asaf Ben Shoham, Abraham Wolde-Tsadick and Karen Lasky (2010)
- No. 101 *TRENDS IN SOUTH AFRICAN INCOME DISTRIBUTION AND POVERTY SINCE THE FALL OF APARTHEID* Murray Leibbrandt, Ingrid Woolard, Arden Finn and Jonathan Argent (2010)
- No. 100 *MINIMUM-INCOME BENEFITS IN OECD COUNTRIES: POLICY DESIGN, EFFECTIVENESS AND CHALLENGES* Herwig Immervoll (2009)

A full list of Social, Employment and Migration Working Papers is available at [www.oecd.org/els/workingpapers](http://www.oecd.org/els/workingpapers).

Other series of working papers available from the OECD include: [OECD Health Working Papers](#).

**RECENT RELATED OECD PUBLICATIONS:**

*INVESTING IN YOUTH: BRAZIL* (2014), [www.oecd.org/employment/action-plan-youth.htm](http://www.oecd.org/employment/action-plan-youth.htm)

*AGEING AND EMPLOYMENT POLICIES: NETHERLANDS 2014*, [www.oecd.org/els/employment/olderworkers](http://www.oecd.org/els/employment/olderworkers)

*SOCIETY AT A GLANCE 2014*, [www.oecd.org/els/societyataglance.htm](http://www.oecd.org/els/societyataglance.htm)

*MENTAL HEALTH AND WORK: UNITED KINGDOM* (2014), [www.oecd.org/els/emp/mentalhealthandwork-unitedkingdom.htm](http://www.oecd.org/els/emp/mentalhealthandwork-unitedkingdom.htm)

*VIEILLISSEMENT ET POLITIQUES DE L'EMPLOI : FRANCE 2014: MIEUX TRAVAILLER AVEC L'ÂGE*, [www.oecd.org/fr/emploi/emp/vieillissementetpolitiquesdelemploi.htm](http://www.oecd.org/fr/emploi/emp/vieillissementetpolitiquesdelemploi.htm)

*MENTAL HEALTH AND WORK: SWITZERLAND* (2014), [www.oecd.org/els/emp/mentalhealthandwork-switzerland.htm](http://www.oecd.org/els/emp/mentalhealthandwork-switzerland.htm)

*PENSIONS AT A GLANCE 2013*, [www.oecd.org/els/public-pensions/pensionsataglance.htm](http://www.oecd.org/els/public-pensions/pensionsataglance.htm)

*HEALTH AT A GLANCE 2013: OECD INDICATORS*, [www.oecd.org/health/health-systems/health-at-a-glance.htm](http://www.oecd.org/health/health-systems/health-at-a-glance.htm)

*OECD EMPLOYMENT OUTLOOK 2013*, [www.oecd.org/els/emp/oecdemploymentoutlook.htm](http://www.oecd.org/els/emp/oecdemploymentoutlook.htm)

*CLOSING THE GENDER GAP: ACT NOW*, [www.oecd.org/gender/closingthegap.htm](http://www.oecd.org/gender/closingthegap.htm)

*OECD PENSIONS OUTLOOK 2012*, [www.oecd.org/finance/privatepensions/](http://www.oecd.org/finance/privatepensions/)

*INTERNATIONAL MIGRATION OUTLOOK 2012*, [www.oecd.org/els/internationalmigrationpoliciesanddata/](http://www.oecd.org/els/internationalmigrationpoliciesanddata/)

*OECD EMPLOYMENT OUTLOOK 2012*, [www.oecd.org/employment/employmentpoliciesanddata](http://www.oecd.org/employment/employmentpoliciesanddata)

*SICK ON THE JOB: Myths and Realities about Mental Health and Work* (2011), [www.oecd.org/els/disability](http://www.oecd.org/els/disability)

*DIVIDED WE STAND: Why Inequality Keeps Rising* (2011), [www.oecd.org/els/social/inequality](http://www.oecd.org/els/social/inequality)

*EQUAL OPPORTUNITIES? The Labour Market Integration of the Children of Immigrants* (2010), via [OECD Bookshop](http://OECDBookshop)

*OECD REVIEWS OF LABOUR MARKET AND SOCIAL POLICIES: ESTONIA* (2010), [www.oecd.org/els/estonia2010](http://www.oecd.org/els/estonia2010)

*JOBS FOR YOUTH: GREECE* (2010), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: DENMARK* (2010), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*OECD REVIEWS OF LABOUR MARKET AND SOCIAL POLICIES: ISRAEL* (2010), [www.oecd.org/els/israel2010](http://www.oecd.org/els/israel2010)

*JOBS FOR YOUTH: UNITED STATES* (2009), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: POLAND* (2009), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*OECD EMPLOYMENT OUTLOOK: Tackling the Jobs Crisis* (2009), [www.oecd.org/els/employmentpoliciesanddata/](http://www.oecd.org/els/employmentpoliciesanddata/)

*DOING BETTER FOR CHILDREN* (2009), [www.oecd.org/els/social/childwellbeing](http://www.oecd.org/els/social/childwellbeing)

*SOCIETY AT A GLANCE – ASIA/PACIFIC EDITION* (2009), [www.oecd.org/els/social/indicators/asia](http://www.oecd.org/els/social/indicators/asia)

*OECD REVIEWS OF LABOUR MARKET AND SOCIAL POLICIES: SLOVENIA* (2009), [www.oecd.org/els/slovenia2009](http://www.oecd.org/els/slovenia2009)

*INTERNATIONAL MIGRATION OUTLOOK: SOPEMI* (2010) [www.oecd.org/els/migration/imo](http://www.oecd.org/els/migration/imo)

*PENSIONS AT A GLANCE 2009: Retirement-Income Systems in OECD Countries* (2009),  
[www.oecd.org/els/social/pensions/PAG](http://www.oecd.org/els/social/pensions/PAG)

*JOBS FOR YOUTH: FRANCE* (2009), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*SOCIETY AT A GLANCE 2009 – OECD Social Indicators* (2009), [www.oecd.org/els/social/indicators/SAG](http://www.oecd.org/els/social/indicators/SAG)

*JOBS FOR YOUTH: AUSTRALIA* (2009), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*OECD REVIEWS OF LABOUR MARKET AND SOCIAL POLICIES: CHILE* (2009), [www.oecd.org/els/chile2009](http://www.oecd.org/els/chile2009)

*PENSIONS AT A GLANCE – SPECIAL EDITION: ASIA/PACIFIC* (2009), [www.oecd.org/els/social/pensions/PAG](http://www.oecd.org/els/social/pensions/PAG)

*SICKNESS, DISABILITY AND WORK: BREAKING THE BARRIERS (VOL. 3) – DENMARK, FINLAND, IRELAND  
AND THE NETHERLANDS* (2008), [www.oecd.org/els/disability](http://www.oecd.org/els/disability)

*GROWING UNEQUAL? Income Distribution and Poverty in OECD Countries* (2008), [www.oecd.org/els/social/inequality](http://www.oecd.org/els/social/inequality)

*JOBS FOR YOUTH: JAPAN* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: NORWAY* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: UNITED KINGDOM* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: CANADA* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: NEW ZEALAND* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

*JOBS FOR YOUTH: NETHERLANDS* (2008), [www.oecd.org/employment/youth](http://www.oecd.org/employment/youth)

For a full list, consult the OECD online Bookshop at [www.oecd.org/bookshop](http://www.oecd.org/bookshop)