METHOD GUIDE 4

Adopting and adapting a standardised modular survey

Kjartan Ólafsson

University of Akureyri, Iceland

November 2016

GLOBAL KIDS ONLINE

www.globalkidsonline.net

Table of Contents

GLOBAL KIDS ONLINE

Global Kids Online is an international research project that aims to contribute to gathering rigorous cross-national evidence on children's online risks, opportunities and rights by creating a global network of researchers and experts and by developing a toolkit as a flexible new resource for researchers around the world.

The aim is to gain a deeper understanding of children's digital experiences that is attuned to their individual and contextual diversities and sensitive to cross-national differences, similarities, and specificities. The project was funded by UNICEF and WePROTECT Global Alliance and jointly coordinated by researchers at the London School of Economics and Political Science (LSE), the UNICEF Office of Research-Innocenti, and the EU Kids Online network.

You can find out more about the author of the report here: www.globalkidsonline.net/olafsson
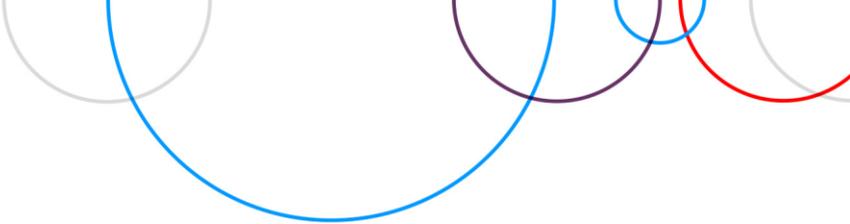
ABSTRACT

Survey design involves a range of different decisions, many of which affect the accuracy of the results. This report discusses some of the key challenges of comparative survey research, and the different approaches to quality in comparative survey projects through the concept of equivalence.

This field of research has developed considerably in the past three decades or so, and we now have a greater understanding of how equivalence can be achieved. Global Kids Online (GKO) has developed a modular survey for those who want to study children's use of digital media. The survey is responsive to local contexts while also allowing cross-national comparisons, and key to its flexibility is the concept of careful adaptation.
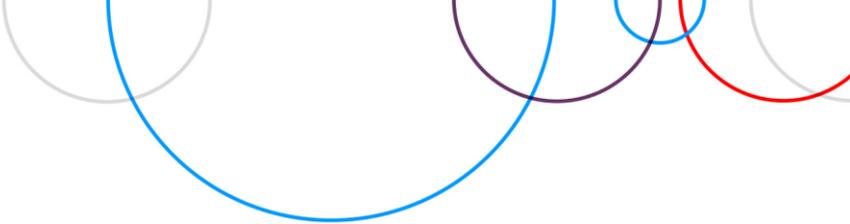
INTRODUCTION

This report discusses the adaptation of survey items (such as those proposed by the Global Kids Online [GKO] project) for a new survey so that they are responsive to diverse circumstances while, where possible, still generating cross-nationally comparable findings. The modular survey that is part of the GKO toolkit is itself adapted from the EU Kids Online survey (carried out in 25 European countries in 2010 – see www.eukidsonline.net) along with a range of items adapted from other surveys. As suggested by the heading of this Methodological Guide, the key to success when constructing a new survey in this way is the concept of adaptation. Adaptation acknowledges that survey questions are meaningful in a particular context that varies over time and between countries (or even areas within the same country). Successful adaptation has to be judged against the type and level of comparability that is aimed for, and the process is likely to involve compromises between different comparability goals.

Survey design and quantitative data collection requires many different decisions that may have serious implications for the eventual results. Compared with the exact science of data analysis and hypothesis testing, survey design can appear as a bit of a dark art, where decisions have to be made with limited information. In his seminal book on question design, Stanley L. Payne (1951) pointed out how researchers often go to great lengths to adjust the finer details of their statistical analysis while ignoring substantial errors caused by question design.

KEY ISSUES

The rise of comparative survey research

The field of comparative survey research has developed considerably since the early 1980s when the first large-scale cross-national studies started to emerge,[1] and many notable research projects have devoted considerable time and energy to improving methods for survey development. Two are of special interest for the current topic of children and media, and are used as examples in this report. One is the Health Behaviour in School-aged Children (HBSC) study (see www.hbsc.org). This was one of the first international surveys on adolescent health, and serves as a useful model in terms of general methodology. Many of the measurements used in various surveys on children's use of media have been adapted from this study. Another notable project is the European Social Survey (ESS) project, which has been at the forefront in developing methods for cross-national comparative research since its inception in 2001 (see www.europeansocialsurvey.org).

> ## "Survey design and quantitative data collection requires many different decisions that may have serious implications for the eventual results."

Judging quality in comparative survey research

Lynn (2003) identifies five broad approaches for judging the quality of cross-national surveys, and most of these also apply in surveys involving comparison over time:

- *The maximum quality approach* aims to achieve a survey of the best possible quality in each country (or at any point in time). This might, however, lead to inconsistencies between countries in coverage, response rate, data collection mode, level of random error, bias and so forth.
- *The consistent quality approach* aims to eliminate between-country inconsistencies by essentially going for the lowest common denominator. Part of this approach might be to let the same organisation carry out the survey in all participating countries – if the countries being surveyed have similar characteristics, this may be a good approach. However, this approach is not suitable for making comparisons between very different countries or over a long period of time.
- *The constrained quality approach* might be seen as a compromise between the two extremes of maximum quality and maximum consistency. The idea is to identify key aspects of the survey design that are likely to affect comparability. These aspects are then constrained in a consistent way between participating countries or in consecutive surveys.
- *The target quality approach* is similar to the consistent quality approach, but aims to set common quality standards in all aspects of the survey at the highest level obtainable in any of the participating countries (this approach cannot effectively be applied to consecutive surveys over time). The idea is to encourage countries to achieve higher standards than they might otherwise have done. However, there is a risk that the end result will contain inconsistencies between countries.
- *The constrained target quality approach* is a modification of the target quality approach: it defines a few key constraints that can also be thought of as entry criteria or minimum standards. The idea is to obtain the advantages of the target quality approach while ensuring consistency on key dimensions. The ESS is an example of this.

The different ways in which survey quality can be approached (Lynn, 2003), and the idea of thinking about survey quality in comparative surveys in terms of equivalence (van de Vijver & Leung, 2011) have

---

[1] For example, the Health Behaviour in School-aged Children (HBSC) study began in 1983, the World Values Survey in 1981, and the International Social Survey Programme in 1984.

important implications for researchers working on survey design. Most importantly, the best results (in terms of survey quality) in studies involving comparison over time or across countries cannot be achieved through simply observing general rules. Equivalence depends on understanding the social and cultural context of the survey, and calls for active collaboration of the individuals involved in its design and implementation.

Criteria for good questions

Theories of the process of answering questions, and the various rules for the design of good questionnaires, have developed significantly since the mid-20th century (see, for example, Bradburn et al., 2004; Saris & Gallhofer, 2007; Tourangeau et al., 2000). However, the main body of research has been developed in surveys for adult populations, and in studies that do not involve comparisons over time or across countries. This points to three important issues facing researchers who wish to design surveys on children and media that can be used for comparisons over time or between countries.

The first key issue is the age of the respondents. Krosnick's (1991) theory of satisficing provides a useful starting point in the quest for improved survey design. The satisficing theory is an extension of the classical model of the question-answering process (Tourangeau & Rasinski, 1988) that identifies four steps when respondents answer questions: (1) understanding and interpreting the question; (2) retrieving information from memory; (3) making a summarised judgement; and (4) reporting this judgement. The satisficing theory distinguishes between two approaches in the question-answering process. The first, 'optimising', is when the respondent consciously goes through all four stages needed to answer a survey question. The second, 'satisficing', is when the respondent gives a more or less superficial answer. From the researcher's point of view, ideally all respondents would use the optimising approach, but satisficing can be related to three dimensions in the question-answering process:

- motivation of the respondent
- difficulty of the questions
- cognitive abilities of the respondent, which is of special importance in studies involving children.

When surveying children, it is therefore necessary to take into account the fact that children do not have the same cognitive functioning as adults. Questionnaires intended for children have to be adapted to the age group for which they will be used. It must also be kept in mind that, although children of a certain age might be able to answer a particular question, it might be ethically unacceptable to ask them to do so.

The second key issue is comparison over time. It is a popular idea that research projects using repeated surveys as a method for measuring social change should aim to minimise changes in the research design. Duncan (1969) laid down this principle in simple terms by pointing out that 'if you want to measure change, don't change the measure'. But it may prove difficult to adhere to this principle when studying constantly evolving media. If studying media is 'to some extent about studying a moving target' (Livingstone, 1998, p. 437), the principle of not changing measurements between consecutive surveys becomes difficult to uphold: this is perhaps one of the reasons why longitudinal designs are so little used for media research. This challenge is likely to increase when the time span of a research project is extended. In such a case, the ideal of standardisation will eventually conflict with the need to collect meaningful information from the respondents or participants in the study.

The third key issue is comparison between countries or across different cultural contexts. Looking beyond national borders for comparative purposes has a long tradition in social science research, but only in the last couple of decades has cross-national (or cross-cultural) comparative research really gained popularity (Hoffmeyer-Zlotnik & Harkness 2005; Smith et al., 2011). Several processes have contributed to this trend: gradual internationalisation of the academic community, removal of political barriers and computerisation of communication. Traditional boundaries – geographical as well as social and cultural ones – are far more easily crossed now. Funding bodies and policy-makers have also been increasingly calling for comparative research, and this call seems to be readily accepted by researchers who find themselves initiating or invited to collaborate in multinational comparative projects (Livingstone, 2003).

Case study: Adapting the EU Kids Online survey in Brazil

In 2012, Brazil was the first Latin American country to adapt the EU Kids Online survey. The translation into Portuguese combined the versions used in Portugal and the UK. Cognitive testing revealed critical points: the wording of the questions was too long for children and too formal for parents; internet-related terms such as 'social network', 'chat room', 'risk' or 'safety' were not understood; and children from less privileged families were particularly tired by the length of the interview and had difficulties reading the self-completion questionnaire. We realised that differences in socioeconomic status were more critical than age differences. This information was valuable for improving the final questionnaire and for a critical reading of the results.

In one of the first questions in the Brazilian questionnaire, TIC Kids Online (2012) invited the children to identify which devices they used for the internet. Adopting and adapting the EU Kids Online questionnaire, the list included: your own desktop; your own laptop or one that you can use in your bedroom; a desktop PC shared with your family; a laptop shared with your family; a mobile phone; TV; tablet; other. The results showed that 37% had selected 'other', thus reporting their access through devices placed outside the home, particularly in LAN[2] houses. The rate of accessing the internet in LAN houses dropped from 35% in 2012 to 22% in 2013 and 2014, thus pointing to a stable use. The question on the devices used was reframed in subsequent surveys.

Case study: Adapting the GKO survey in the Philippines

The major challenge in adapting the GKO survey to the Philippine context was related to the large number of spoken languages and dialects in the country. The population speaks between 120 and 175 languages, depending on the method of classification, and two languages are considered

official – English and Filipino (the standard form of Tagalog language). The pilot research took part in Manila where the spoken languages are Filipino/Tagalog and English, and in the province of Pampanga, where the population also speaks Kapampangan. Aiming to make the survey more accessible and closer to the children's everyday spoken language, the research team translated the English version of the survey into Filipino/Tagalog, and used this as the main language of the survey. Some immediate difficulties became apparent in translating from English, for example, in relation to the 'yes' and 'no' response options, as there are two ways of responding 'no' and two ways of responding 'yes' in Tagalog. In addition, some children in the province of Pampanga knew the official Tagalog language but mainly spoke Kapampangan in their everyday life or used the informal version of Tagalog. They struggled with some of the more complex phrases in the survey describing the language as 'deep Tagalog' (formal Tagalog). Many, including children with both public and private school education, said they would rather have done the survey in English. Adapting the survey and administering it in such a multi-lingual environment thus had to accommodate both the complex daily language practices of the children and varying language proficiencies. In this particular case, these issues were handled by recruiting researchers from the local community who were prepared to offer explanations and clarifications in any of the local languages.

Case study: Adapting the survey to the local context in Serbia

In Serbia, the debate around online piracy, that is, downloading copyright-protected content, is featuring prominently in media and policy discourses. However, the GKO survey does not contain any questions about downloading copyright-protected content, partly because asking such questions in certain countries might put the child respondent at risk, since enumerators could be obliged to report such criminal offences to the authorities. Because online piracy was a key issue of interest in Serbia, the Serbian team added a

---

[2] A LAN house is a business where a computer connected over a Local Area Network (LAN) to other computers can be used, often for the purpose of playing multi-player computer games.

number of questions on online piracy to the child and parent questionnaires. The GKO survey has been designed to allow for such adaptation to national or cultural settings, and facilitates the incorporation of questions that might only be of relevance to a few countries. The questions on online piracy, developed and subsequently tested by the Serbian team, will not be included in the core part of the GKO survey, but the scale and its psychometric properties will be included in an appendix and made available for other researchers, should they find it relevant to include in a future country study.

> **"In studies involving comparison between countries or over time, the research design aims to limit both random error and systematic error."**

Equivalence

In studies involving comparison between countries or over time, the research design aims to limit both random error and systematic error (or bias; see below). In comparative research the overall goal of limiting errors has to be obtained while at the same time striving for equivalence (van de Vijver & Leung, 2011). 'Equivalence' broadly means that in all aspects the survey as a measurement tool works in the same way across countries or over time. The goal of equivalence applies to all aspects of the research design, and has important implications for decisions made in the design process. It also applies to the data collection process and the ways in which fieldwork is conducted.

Errors in survey research

All numbers obtained through surveys contain two error components, random error and systematic error (or bias); the goal throughout the survey process is to limit and control these. Random error is present in all data obtained by drawing a sample from a population (as is almost always the case in surveys), and is simply the (random) difference between sample and population values. The random error is simple to deal with as it is directly related to the sample size and can be reduced by increasing the effective sample size.

The random error can also be easily estimated and is what is being controlled for in significance testing. However, systematic error (or bias) is more difficult to deal with, as it stems from the research design and is the result of several factors that may be difficult to estimate.

The various sources of systematic error mainly fall into two categories: errors resulting from the sampling and data collection procedure, and errors resulting from measurements and data processing.

An example of sampling error would be if certain groups in the population were less likely to end up in the sample. Online surveys, for example, include only those who have access to the internet – everyone else is unable to participate. In a cross-national comparative study of countries with different levels of internet penetration, this would result in between-country variation that could be partly caused by difference in selection bias between the countries.

> **"An example of sampling error would be if certain groups in the population were less likely to end up in the sample."**

Measurement error resulting from question design is probably the most easily identified source of systematic error, and research projects focusing on comparison between countries or over time will invariably put serious effort into limiting this error. As studies grow in size (be it in terms of the number of countries involved, the number of data collection rounds in a long-term study or just in terms of the number of people involved in the data analysis), it becomes increasingly important to have clear procedures for coding and data handling. This applies, for example, to the treatment of missing values and procedures on data cleaning.

> **"'Equivalence' broadly means that in all aspects the survey as a measurement tool works in the same way across countries or over time."**

As outlined above, errors caused by the research design are most frequently related to systematic error
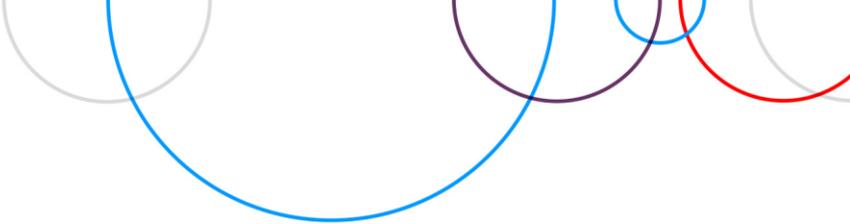
(bias) in point estimates. However, it is also important to keep in mind that the research design (in particular, the design of measurements) can result in random measurement error (Saris & Gallhofer, 2007). This does not affect point estimates (as this kind of error results in random fluctuations around the true value), but it can seriously limit the strength of correlations observed in the data. This type of random error might be seen as equivalent to 'white noise' in audio, which obscures the music or the spoken word. In survey research, random measurement error reduces the observed strength of correlations, leading researchers to underestimate the importance of variables in the analysis.

MAIN APPROACHES

The importance of adaptation

The key to a successful outcome is the concept of adaptation itself. Successful adaptation has to be judged against the level or type of comparability that is aimed for. The adaptation process is therefore likely to be one of compromises between different goals in terms of comparability. In many cases, the aim will be to make comparisons on the level of so-called point estimates (e.g., percentage of children engaged in a certain type of activity). However, this kind of comparison calls for the highest level of standardisation in sampling and measurement, because any systematic bias in the measurements will directly influence such comparisons.

> ## "Successful adaptation has to be judged against the level or type of comparability that is aimed for."

It is therefore worth bearing in mind that it is possible to make comparisons in terms of correlations (e.g., the effect of gender on whether children engage in a certain type of activity). Such comparisons can often be made even if it is not possible to compare the point estimates directly. For example, two studies might ask how frequently children engage in an online activity, but the response scales might be so different that direct comparison of frequency between the two studies is not feasible. However, it might be possible to directly compare the effect of gender on how frequently children engage in that activity between the two studies.

Tensions in comparative surveys

In their discussion on the development of the HBSC study, Roberts and colleagues (2009) identified four areas of tension that arise in a project carrying out repeated surveys. The same tensions are likely to arise in any survey based on previous data collection efforts. Expanding the analysis of tensions in the HBSC study (Roberts et al., 2009) to research projects wanting to build on the EU Kids Online survey in a

wider context may be helpful in identifying potential challenges and solutions.

> ## "Comparative survey projects tend to grow to cover more countries and more topics."

The first tension is maintaining quality against the background of growth. Comparative survey projects tend to grow to cover more countries and more topics. The EU Kids Online survey of 2010 was developed for implementation in 25 European countries at a given point in time. The questionnaire and methods were designed to meet the specific demands and affordances of that occasion. The theoretical model underpinning the survey has since been revised, and more countries have joined the network. This can lead to difficulties in maintaining comparability while also respecting the needs of individual countries and researchers who might wish to ask questions not included in the previous survey.

To solve this dilemma, many comparative research projects define a set of core questions that become the basis for comparison between countries or over time, and that can then be combined with optional and country-specific items. Obviously the choice of core items is of great importance, as these become the basis for comparisons; for this reason, it is also important that the core items are included in as many surveys as possible. In fact, many comparative projects make all core items mandatory for all participating partners.

> ## "Many comparative research projects define a set of core questions that become the basis for comparison between countries or over time."

There can, however, be circumstances where core questions have to be omitted from a survey, such as when answering a question might be perceived to incriminate a respondent (e.g., children admitting illegal conduct carrying severe punishment), which put
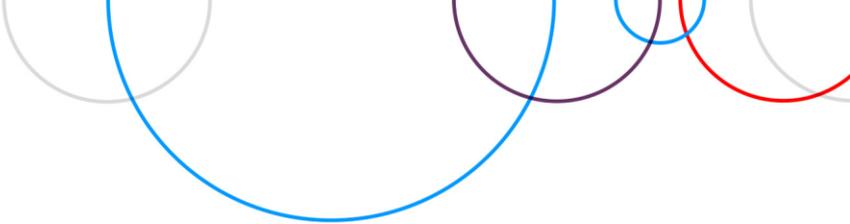
interviewers in a position where they have to break their confidentiality (e.g., if there is a requirement to report certain behaviour) or when the inclusion of a topic is likely to have negative consequences for the whole survey process (e.g., where it might be seen as inappropriate to ask children about a certain topic, which might lead to lower response rates).

"The key to a successful outcome is the concept of adaptation itself."

The second tension is improving quality despite restricted financial resources, although this challenge is, of course, not limited to comparative research projects. Developing a survey from existing modules makes a good starting point, but sufficient resources should be set aside for the adaptation process – the resources saved in not having to design a survey from scratch have to be used instead for the adaptation. In cross-national comparative projects, countries will invariably have different resources at their disposal. Countries with limited resources may have to choose between the activities that might improve quality (sending researchers to international meetings, testing survey items beyond normal pilot testing, training staff, etc.). It is possible to set up a system of direct transfer of funds between partners, but this is not always easy to operate in practice. Alternatively, some partners in a collaborative project can take responsibility for central tasks (such as questionnaire development or database management), to the benefit of all partners.

"It is possible to set up a system of direct transfer of funds between partners, but this is not always easy to operate in practice."

The third tension is between monitoring trends and improving or adapting questionnaire content between consecutive surveys. Duncan's (1969) principle ('if you want to measure change you shouldn't change the measure') creates tension when a survey needs to be adapted to capture changes in the phenomena being studied, and also when efforts are made to improve the survey design. In a research design that uses the idea of core questions, these provide the basis for comparison between surveys. But this puts important

constraints on how the core questions can be developed and altered, as outlined by Lynn (2003) in the five approaches to setting quality standards in surveys. For any given item that has been identified as a core item in a comparative survey, it is usually possible to suggest an alternative item or a change to the original item that would work better in a particular context. With the consistent quality approach, no changes would be considered, but thinking back to Lynn's (2003) five approaches (see Section 1.2), it cannot be seen as the only alternative.

"While recognising the cultural divide between researchers and policy-makers, it is also important to recognise the positive potential of genuine collaboration between these important groups."

The fourth tension is between the requirements of scientific and policy audiences. There has recently been increased dialogue between researchers and policy-makers, which many would say has benefited both parties. While recognising the cultural divide between researchers and policy-makers, it is also important to recognise the positive potential of genuine collaboration between these important groups. Some research projects have sought to formalise the dialogue between research and policy, and to engage in an active dialogue with policy-makers in the design process. Such collaboration can take place both at a national and international level.

IDENTIFYING GOOD PRACTICE

Focus on the theory

A clear theoretical framework is essential to designing a good survey. Approaching the survey design from the viewpoint of a theoretical model helps in defining research questions, formulating hypotheses and setting out a clear path from concept to operation. Adapting an existing modular survey for use in a new study is thus not only a matter of revising a set of questions or of adding new questions; it is also a matter of developing a theoretical framework for that study (even if it is just adding new countries or new waves of data in an existing research project). Having a well-defined theoretical framework in a study is of great value when key decisions about priorities in the research design are being made; in a comparative study it will include also the comparative element. A comparative study would not include comparison just for the sake of comparison.

"A clear theoretical framework is essential to designing a good survey."

The theoretical model is also the key to selecting the most relevant concepts and the best items to measure those concepts. Gone are the times when the length of a questionnaire was limited by the size of punch cards or computer memory, but the time a respondent is willing to spend answering a survey is limited, and puts limitations on the number of questions in a survey. Allowing a group of researchers to suggest questions they would like to see included in a survey will almost inevitably result in an over-long list of questions. The bigger the group of researchers, the longer the list of interesting questions is likely to become. However, items should not be included in a survey because they are interesting but because of their role in the data analysis. When a questionnaire is too long it is possible to use tricks such as split-half designs (where not all respondents answer all the questions), but this will increase the cost of data collection and complicate the data analysis.

A modular approach

A modular approach to survey design is intended to help maintain a focus on the theoretical model. The idea is that for each part of the theoretical framework there will be a survey module; within each module a range of questions allows researchers to examine a particular topic. Often the items are divided into 'core' (mandatory items in the HBSC study) and 'optional' items. In comparative studies where many researchers are involved, ideas for new survey modules will almost certainly arise. The ESS has an open call for proposals for new modules before each new round of the survey. The HBSC study has a set of guidelines for those wanting to promote new items or survey modules. Having clear procedures on how to propose new modules should encourage suggestions on how to develop the study further. Such proposals would, of course, have to meet the quality standards of the survey.

When the outcome of the survey design process is to change survey items that have been used previously, it is always desirable to estimate the impact of such changes. Ideally this would be part of the design process (e.g., estimating the effect of a change in question wording on point estimates before deciding whether to take up the new wording). The HBSC project has set standards for changes in mandatory items (changing existing items or adding new ones): it takes a minimum of eight years to implement such changes (Roberts et al., 2009). Such a strict approach would hardly be feasible in a study focusing on children and media, because of the fast-changing nature of the field, but some kind of testing of the possible effect of changes in the research design is necessary.

"The bigger the group of researchers, the longer the list of interesting questions is likely to become."

Translation

A common approach in many cross-national survey projects is to design a questionnaire in one language and then to translate or adapt it for use in other languages. The ESS project has developed a clear framework and guidelines for this part of the adaptation process. A key feature of the process is the TRAPD (translation, review, adjudication, pre-testing and documentation) methodology. The process is set out in the translation guidelines available in the methods section of the project website (see www.europeansocialsurvey.org/methodology/translation.html), where there is also a translation quality checklist.

> **"A common approach in many cross-national survey projects is to design a questionnaire in one language and then to translate or adapt it for use in other languages."**

The ESS puts a selection of key items through two additional steps, verification and Survey quality Predictor (SQP) coding. The verification process involves a linguistic quality check (outlined in detail in the methods section of the project website). The SQP is a survey quality prediction system for questions used in survey research (see http://sqp.upf.edu/). The program predicts the reliability and validity of the questions based on their formal characteristics. Predictions are based on a meta-analysis of a large number of multi-trait multi-method (MTMM) experiments containing 4,000 items in around 20 languages.

Coding

A coding scheme for any survey will usually be set out in a data dictionary where the exact coding of all questions is specified. For data files it is important to use clear labelling and to stay as close as possible to the original wording of questions to prevent concept drifting. If, for example, respondents are asked if they have been treated in a hurtful or nasty way, this phrasing should be used in the variable label rather than using terms such as 'bullying' to remind those

using the data that this is how bullying was operationalised in the survey.

Analysis of item non-response is an important part of quality control of survey items, and consistent use of missing values in the questionnaire and other documentation ensures that skips in the interview are coded consistently. It is therefore of value to use separate missing values to denote different types of missing values. The ESS, for example, distinguishes between four types of missing values (see Norwegian Social Science Data Services, 2015):

- Not applicable (possibly coded as 6, 66, and 666)
- Refusals (possibly coded as 7, 77, and 777)
- Don't know (possibly coded as 8, 88, and 888)
- No answer (possibly coded as 9, 99, and 999).

In some cases, an internationally recognised coding scheme can be used: for education there is the International Standard Classification of Education (ISCED, see www.uis.unesco.org); and for occupation there is the International Standard Classification of Occupations (ISCO, see www.ilo.org). It is also recommended to use standard abbreviations when these are available as, for example, for country names (ISO 3166 defines codes for the names of countries, dependent territories, special areas of geographical interest and their principal subdivisions). Standard abbreviations are also available for languages (ISO 639).

> **"A coding scheme for any survey will usually be set out in a data dictionary where the exact coding of all questions is specified."**

Missing values and routing

Routing can be used to allow respondents to skip questions that do not apply to them. For example, respondents who have not seen any sexual images in the past 12 months should be allowed to skip questions that depend on having seen such images. Great care has to be taken when routing is implemented so that respondents are not accidentally routed away from questions to which they should have responded (this is particularly important when making

use of data collection software or computer-assisted personal interviewing [CAPI]).

> ## "Great care has to be taken when routing is implemented so that respondents are not accidentally routed away from questions to which they should have responded."

Routing can also create challenges in data analysis. Using sexual images again, as an example, if respondents who have not seen a sexual image in the past 12 months are routed away from a follow-up question (such as how upset they were by seeing such images), this has to be properly accounted for when reporting the respondents who have been upset by sexual images. To facilitate this, it is important that respondents who have been routed out in the follow-up question are indicated by a specific missing value.

Cleaning data

Data cleaning broadly refers to corrections and adjustments that can be made during and after data entry. Such corrections include correcting mistakes in the entry of data (when a response to a question has been entered incorrectly) and solving inconsistencies between different variables (e.g., when a respondent is coded as having not seen any sexual images in the past 12 months and then in a different variable as having been 'just a little upset' by seeing such images online). The following two tables show some examples from the data editing and cleaning process of EU Kids Online and South African Kids Online, respectively.

Table 1: Example data edits carried out in EU Kids Online

Questions	Approach and edits applied
Child age: Checking contact sheet: SCR.3b/4b Age of selected child against the child age question in the parent questionnaire: Q.201 What is the age of your child?	The age of the interviewed child in the contact sheet was edited, where necessary, to ensure it referenced the child who had completed the questionnaires. If there was more than one possible match (among the children recorded in the contact sheet data) then the child that uses the internet (SCR3D) was identified as the selected child. If both/all (or neither/none) used the internet then one child was selected at random. In order to avoid confusion, the contact sheet selected child age variable was not included in the main survey data set (just in the contact sheet data set). This ensured that all data users will use the same variable for analysis on child age (as recorded during the main interview). All selected children were then coded as internet users at SCR3D for consistency (as per the profile of survey participants desired).
Child gender: Checking contact sheet: SCR.3c/4c Gender of selected child against the child gender question in the parent questionnaire: Q.201b Gender1 of child?	As above
Number of children living in house: Checking contact sheet: SCR.2 Number of children aged 9-16 living in the household against parent questionnaire variable: Q202 number of children aged	If more children were reported at SCR2 than Q202, Q202 was edited to be equal to the response at SCR2. If there was no valid response at Q202 and SCR2, answers were back- coded from SCR3. If there was no data recorded at SCR2, SCR3 and Q202 responses were edited to refer to 1

0-17 living in the household.

child.

Child use of communication media on the internet: Q324a-f asked children which of a range of activities they had done in the last year. This was checked against answers at Q308a-f which asked how often they had done the same activities in the past month.

- email usage
- visited a social networking profile
- Visited a chat room
- used instant messaging
- Played games with other people on the internet
- Spent time in a virtual world

If a child had coded "no" (not done in the past year) at Q324 for activities they had reported doing in the past month at Q308, the response at Q324 was edited to show that they had participated in it.

Table 2: Example data cleaning carried out in South African Kids Online

Issues	Approach and edits applied
Data collection and capturing processes	The South African Kids Online data collection and capturing processes involved multiple stages of quality controlling and data cleaning. This involved enumerators checking the quality of their questionnaires as they completed them, as well as their supervisors checking the quality of each questionnaire. Completed questionnaires were returned to the CJCP office and checked for quality by the researcher working on the project. The questionnaires were then captured on Statistical Package for the Social Sciences (SPSS) by data capturers who were trained on how to correctly capture the questionnaire and what errors to look out for when capturing.
Screening and validating	Once capturing was complete, the data was manually checked for any errors by the researcher. The CJCP data validation process generally consists of two parts – the first, involving the screening of all of the data to identify any questionable or erroneous values in the data, and the second, involving the individual assessment of cases to decide on the appropriate response to the questionable values identified in the first stage. Possible responses would include retaining the value, rejecting it as invalid or replacing the questionable value with a "missing" value. Throughout the cleaning process our main focus was checking for consistency within the data set.
Cleaning	Within our data set, cleaning involved checking for missing data, which was missing as a result of capturing error, or as a result of missing data in the paper questionnaire itself. We did this by checking the frequencies of each item. Cleaning also involved cross-tabing variables that related to each other to check for consistency. So for example, we checked that those children who said they were in primary school didn't also say that they had a university degree.
Checking	An additional element of cleaning was making sure that skip patterns were correctly followed and

Routing	this proved to be an especially time consuming part of the data cleaning process. The South African team had four different labels that were given to questions and each question needed to be answered by a different group of participants. These were:

- All participants
- Internet users
- All participants aged 12 and older
- Internet users aged 12 and older

So for each of these questions, it was necessary to ensure that the correct group of respondents had answered the questions. Where respondents hadn't answered the questions, even though they should have, the cell was deemed to be missing data. When an enumerator had filled in the response option and they shouldn't have, there responses were removed.

Many questions in the survey also had skip patterns that did not relate to these classification labels. For these skip patterns, the data was also checked for consistency to make sure that the correct people responded to the specific questions.

Identifying non-users	The South African team would recommend that if future studies consider including non-internet users in the sample, that they use an additional non-internet user module rather than having the internet user and non-internet user questions integrated, which made data collection, capturing and cleaning unnecessarily difficult. In general, we would advocate for only including skip patterns where absolutely essential, rather than just for 'nice to have' purposes.

Source: Centre for Justice and Crime Prevention, South Africa

"Data cleaning broadly refers to corrections and adjustments that can be made during and after data entry."

For cross-national or long-term research projects it is obviously important that data cleaning is done consistently between countries and over time. Sometimes the data dictionary is used to keep track of decisions made on data cleaning, as these will relate to individual questions.

Taking care of the data during analysis

In every research project, the first steps in data analysis are important. Based on the theoretical model behind the research, key concepts will be reflected in the questionnaire. The key concepts are often measured in a set of variables that then are added up– these sums result in a range of derived variables that are used in the analysis. Respondents in a study might, for example, be asked if they participate in certain

online activities; these questions could then be used to form a measurement of the number of activities, which would be a derived variable in the data. It is important to provide clear information on how such derived variables have been obtained, so everyone working with a particular data set can use derived variables in the same way.

"In every research project, the first steps in data analysis are important. Based on the theoretical model behind the research, key concepts will be reflected in the questionnaire."

During the initial analysis it is also possible to discover errors or inconsistencies; it is therefore useful to distinguish between different 'final' versions of the data set. There can be a 'final' version of the data after it has been checked for errors and inconsistencies but before any derived variables have been created. This data set is then final in the sense that it is ready for analysis. Then there can be a 'final' version of the data

after the initial analysis and where important derived variables are included.

Documenting the survey process

Careful documentation of the survey process is crucial in cross-national surveys and in surveys that include repeated rounds of data collection. As a rule of thumb, the greater the number of people involved in a survey (whether in data collection or data analysis), the more important it is that every step of the survey process is carefully documented. For large-scale projects such as the ESS, the sheer volume of such documentation can be almost overwhelming, but it should be remembered that it has two distinct purposes. For those participating in a research project, clear documentation can be used to improve consistency and to prevent tacit knowledge being lost if key people leave the project. Those outside the project might use documentation of the survey process to replicate or evaluate the study.

The documentation should include the following information:

- names, labels and descriptions for variables, response options and their values
- explanation of codes and classification schemes used
- codes of, and reasons for, missing values
- derived data created after collection, with code, algorithm or command file used to create them
- weighting and grossing variables created, and how they should be used.

Sharing data

Research data is a valuable resource that is costly and time-consuming to produce. Survey data can have a significant value beyond the original research, and so there has been an increased emphasis on sharing data. Several data archives offer researchers the possibility of having their data professionally curated so that it becomes easily accessible, both in the short term and in the future. New and innovative research can then be carried out based on existing data, and results can be verified by repeating an analysis.

Most data archives will ensure the following (see, for example, the UK Data Archive):

- safekeeping of research data in a secure

environment
- long-term preservation and back-up of data
- resource discovery of data through inclusion in online catalogues
- rights management of data (licensing issues)
- administration of data access
- enhancing the visibility of data, thus enabling more use and citation
- management and monitoring of data use
- promotion of data to user groups.

If there is a desire for data to be shared, this should be made clear at the outset so that the survey can be designed and prepared with this in mind. Some funding agencies insist on data being made available, which can be a challenge in cross-national surveys: if funding bodies in one country insist on data being made publicly available, this might create challenges for others involved in the research. Data sharing might not be possible unless it has been stated clearly in applications to the relevant ethics committees.

> ## "When data is shared through data archives, certain documentation of the survey process will be required."

When data is shared through data archives, certain documentation of the survey process will be required. The relevant data archives will state clearly the minimum documentation required; in most cases, a variety of additional documentation can accompany the data (see, as an example, the record for the 2010 EU Kids Online survey in the UK data archive at https://discover.ukdataservice.ac.uk/#6885).

If data is to be submitted to a public archive, it will obviously have to be anonymised. Due care must also be taken in the level of detailed information provided in certain variables that might allow individuals to be identified (through a set of variables such as age, gender, municipality, school, parent occupation, etc.).

USEFUL ONLINE RESOURCES

Resources provided by the author

EU Kids Online (2010). *Research toolkit.* London: EU Kids Online, LSE. http://lse.ac.uk/EUKidsOnline/Toolkit

European Social Survey (ESS) (2014). *ESS methodological overview.* www.europeansocialsurvey.org/methodology/

Health Behaviour in School-Aged Children (HBSC) (no date). *Survey methods.* www.hbsc.org/methods/index.html

Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). *Technical report and user guide: The 2010 EU Kids Online survey.* London: EU Kids Online, LSE. http://eprints.lse.ac.uk/45270/

Norwegian Centre for Research Data (NSD): one of the largest archives for research data of its kind, and stores the European Social Survey data. www.nsd.uib.no/nsd/english/

Richardson, D., & Ali, N. (2014). *An evaluation of international surveys of children.* OECD Social, Employment and Migration Working Papers, No. 146, Paris: OECD Publishing. http://dx.doi.org/10.1787/5jxzmjrqvntf-en

UK Data Archive: EU Kids Online 2010 survey, University of Essex (2002–16). www.data-archive.ac.uk/

Universitat Pompeu Fabra (no date). *Survey quality predictor.* http://sqp.upf.edu/

World Values Survey, Wave 6 (2010–2014). Official Aggregate v.20150418. World Values Survey Association. www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Additional resources

Fowler, F. J. (2008). Chapter 6: Designing questions to be good measures. In *Survey research methods.* London: Sage. http://uk.sagepub.com/sites/default/files/upm-binaries/23856_Chapter6.pdf

EU Kids Online: Best Practice Guide. http://www.lse.ac.uk/media@lse/research/EUKidsOnline/BestPracticeGuide/Home.aspx

REFERENCES

Bradburn, N., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design.* San Francisco, CA: Jossey-Bass.

Duncan, O. D. (1969). *Indicators of social change.* New York: Russell Sage Foundation.

Hoffmeyer-Zlotnik, J. H. P., & Harkness, J. (2005). Methodological aspects in cross-national research. *ZUMA-Nachrichten Spezial*, 11.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.

Livingstone, S. (1998). Introduction. *European Journal of Communication*, 13 (4), 435–451.

Livingstone, S. (2003). On the challenges of cross-national comparative media research. *European Journal of communication*, 18 (4), 477–500.

Lynn, P. (2003). Developing quality standards for cross-national survey research: five approaches, *International Journal of Social Research Methodology*, 6 (4), 323–336.

Norwegian Social Science Data Services. (2015). *ESS7 2014 Data protocol, Edition 1.2, April 2015.* Bergen, Norway: Norwegian Social Science Data Services.

Payne, S. (1951). *The art of asking questions.* Princeton, NJ: Princeton University Press.

Roberts, C., et al. (2009). *The Health Behaviour in School-aged Children (HBSC) study: methodological developments and current tensions.* Basel, Switzerland: Birkhauser Verlag.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation and analysis of questionnaires for survey research.* Hoboken, NJ: John Wiley & Sons.

Smith, S. N., Fisher, S. D., & Heath, A. (2011). Opportunities and challenges in the expansion of cross-national survey research. *International Journal of Social Research Methodology*, 14 (6), 485–502.

TIC Kids Online Brasil. (2012). *ICT Kids Online Brazil 2012: Survey on internet use by children in Brazil.* Sao Paolo, Brazil: Cetic.br, http://cetic.br/media/docs/publicacoes/2/tic-kids-online-2012.pdf

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* New York: Cambridge University Press.

van de Vijver, F. J. R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models and data analytic procedures. In D. Matsumoto & F. J. R. van de Vijver (eds) *Crosscultural research methods in psychology* (pp 17–45). New York: Cambridge University Press.

CHECKLIST 1

Glossary of key terms

Term	Description
Equivalence	The extent to which measurements or questions used to capture a certain construct will work in the same way in every country.
Standardisation	When used in statistics, standardisation refers to the transformation of measurement scores into z-scores (transformation of raw scores into standardised units of measurement). In survey design, standardisation is closely related to the concept of equivalence, and refers to the idea of harmonising as many aspects as possible when implementing a survey across countries or over time.
Survey mode	The method of data collection in a survey is referred to as the survey mode. Collecting data face-to-face is one survey mode; collecting data through an online questionnaire is another.
Survey item	Questions in surveys are often referred to as items. A survey item often consists of several questions or response options; it may also be a single question.
Data dictionary	A document listing all questions in the questionnaire with information on names, labels and coding instructions. It may also include information on derived variables, treatment of missing values and rules for data cleaning.
Optimising	When a respondent consciously goes through all four stages needed to answer a survey question: (1) understanding and interpreting the question; (2) retrieving information from memory; (3) making a summarised judgement; and (4) reporting.
Satisficing	When a respondent gives a more or less superficial answer without consciously going through all the steps necessary to give the most accurate answer to a survey question
Point estimates	When samples are used to calculate a value that is intended to serve as the best estimate of a population parameter. For example, when survey data is used to calculate the number of children participating in a certain online
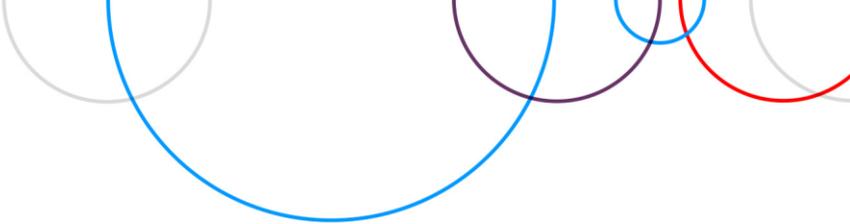
activity, that number is a point estimate.

Random error	The difference between sample and population values. Sometimes called sampling error, it is the random fluctuation in results observed if we were to take repeated samples from the same population. We use significance testing to estimate and control for random error. The random error is directly affected by the sample size: by increasing the sample size fourfold the random error is halved.
Systematic error (bias)	These are errors caused by the sampling procedure (e.g., by particular groups being systematically less likely to be included in the sample) or by the measurements (e.g., by a poorly designed question). The source of these errors is essentially in the design of the study, so these errors are not affected by the sample size.
Floor/ceiling effects	When the response options offered do not fully capture the variability in the attitudes or behaviour of respondents. An example of ceiling effects would be to offer 'daily' as the highest option for an activity engaged in on a daily basis by the majority of respondents.

CHECKLIST 2

Key issues to consider

Term	Description
Survey mode	The survey mode (method of data collection) will be reflected in the question wording and the overall design of the questionnaire. Questions may therefore have to be modified if, for example, a survey designed for use in schools is to be used in face-to-face interviews carried out at home.
Pre-testing and piloting	Where time and resources are limited, it can be tempting to skimp on pre-testing and piloting. However, as these are vital components of the research process, they require due diligence.
Whether to change the questions	A key question when adapting questions for different countries and over time is whether to make changes or not. All changes have the potential to prevent comparisons, so there is a tendency towards conservatism. However, it is possible that not making changes when surveys are repeated over time or adapted between countries may also prevent comparisons being made.
Translating questions	The concept of translating questions is in many ways misleading, and it is important to engage individuals with experience and expertise in survey research in this process.
Countries sharing the 'same' language	Questions developed and used in one country might require adaptation before being used in another country, even if these two countries share a common language.
The use of standardised scales	Using questions that form some kind of a standardised measurement tool (psychometric scales, for example) might seem a safe and straightforward option. However, such tools tend to include many questions and should therefore be used with caution in population surveys.

Response scales	The choice of response categories is often context-specific: a set of response options that works well in one country might result in serious floor/ceiling effects for the measurements when used in a different context. Adaptation of response scales should, therefore, be carefully considered.
Routing	In surveys where not all respondents are expected to answer all questions, routing can be used to allow respondents to skip questions, which, by definition, do not apply to them. This has to be implemented carefully, however, and the feasibility of this approach depends to some extent on the survey mode. Routing also has to be taken into consideration in the coding of the data so that it can be properly accounted for in the data analysis.
Data archiving	There can be many benefits in depositing data to a public archive. However, the decision to make the data publicly available should be made before the data is collected, and should be communicated clearly to everyone involved.