A Pan-European Study on Children's Online experiences: Contributions from Cognitive Testing

Leslie Haddon*, Cristina Ponte**

*LSE, London School of Economics
**FCSH, Universidade Nova de Lisboa

Abstract

The *EU Kids Online* research on children's experience of and ability to cope with risks faced several challenges: the large number of countries involved (25 countries); 2) interviewees were children aged from 9-16 (along with one parent), 3) the main focus was on the sensitive topic of online risks. In order to address these challenges and prior to the survey, cognitive testing was conducted in all the participating countries. This involved in-depth interviews to evaluate children's understanding of the questionnaire, their ability to answer each question, any ambiguities or other difficulties that emerged and in general how their felt about the process of answering the questions. This paper reports the issues that arose, how the questionnaire was (re)designed, and hence the overall relevance of these tests and the insights they provide more generally for cross-national surveys and those aimed at children.

Keywords: cognitive testing, qualitative analysis, survey, mixed methods, EU Kids Online II

1. Introduction

There has been some discussion of the different ways that qualitative and quantitative researched can be combined, and although there are issues in doing so (Bryman, 2006a, 2006b), this can produce a *complementarity of methods* that creates opportunities for a more holistic approach to measurement, analysis and interpretation (Lobe et al, 2008). In drawing up taxonomies of ways of combining these approaches one justification for doing so has been called *development*: 'seeks to use the results of one method to help develop or inform the other method, where development is broadly construed to include sampling and implantation as well as measurement decisions' (Greene et al, 1989, 259). Another more refined taxonomy refers to *instrument development*: 'contexts in which qualitative research is employed to develop questionnaire and scale items – for example, so that better wording or more comprehensive closed answers can be generated' (Bryman, 2006c, 106). This article focuses on just such an, in fact established, form of qualitative analysis that is built into the survey design process itself. The aim is to reflect on the experience of actually utilising the qualitative analysis to achieve the developmental goal, and also to ask whether the term 'instrument development' fully does justice to this process.

Quantitative researchers have for some time employed various forms of pilot study of survey questionnaires in order to gain formal or informal feedback about how the interview process went, interviewee reactions,

problematic sections, etc. Other researchers have been known to try out questionnaires on friends and family and ask for feedback. But an approach developed for quantitative studies called cognitive testing aspires to attain such feedback in a more comprehensive and systematic fashion. Hence the first section examines what cognitive testing aims to do, its different components and goals. But the main interest of this article is in how to make use of this cognitive testing feedback when developing a particular empirical study. Moreover this was an international project that was always going to be demanding because of the existence of different languages and cultural factors that may affect the understanding of the questions being asked.

The EU Kids Online II study, involving a survey in 25 European countries was focused on the risks faced by children when using the internet[1]. In particular, the EU Kids Online II research on children's experience of and ability to cope with risks faced online encountered specific and additional challenges. These related to: 1) the large number of countries involved; 2) the fact that the interviewees were children aged from 9-16 (along with one parent); 3) the fact that the main focus was on the sensitive topic of online risks. This raised issued relating to such matters as the attention span of young children, the sensitive nature of the area being studied, especially for younger children, the types of words they understand compared to adults, and cultural differences in the connotations of the language used.

In order to address these challenges and prior to the survey, cognitive testing was conducted in all the participating countries. This involved in-depth interviews to evaluate children's understanding of the questionnaire, their ability to answer each question, any ambiguities or other difficulties that emerged and in general how their felt about the process of answering the questions. The following sections reports the issues and indeed problems that arose, how the questionnaire was (re)designed more than once because of this, and the limitations of what could be achieved in actual questionnaire design so that some of the insights from the qualitative work in practice have more of a relevance for the interpretation of the survey findings. Ultimately the cognitive testing was useful in a number of ways, but here the interest is in how the wider research process accommodates the feedback from this particular research tool.

1. The Goals and Elements of Cognitive Testing

In media studies, it has been argued that cognitivism represents a 'third' academic culture, in addition to the cultures of social sciences and humanities which have been converging over the past few decades (Jensen, 2002). Since the early days of psychology, the process by which individuals assimilate and organise information about events and relationships has attracted theorists interested in perception,

---

[1] More information at www.eukidsonline.net

memory, thinking and language. Whereas the first major research approach had been one of introspection, with the rise of behaviorism in the 1960s cognition started to be reconsidered as an essential research area where experimental techniques could be applied. This was accompanied by more sophisticated methodologies, including the application of information processing models to cognitive activity, far beyond the early introspective approaches (O'Sullivan et al, 1994).

Arising from these developments in social and cognitive psychology, cognitive testing has been used by social researchers since the 1980s. This testing focuses on the problems related to data collection, frequently those relating to the understanding of and responses to questionnaires. According to the cognitive testing literature (Tourangeau Rips & Rasinki, 2000; Collins, 2003), this approach mainly looks at the *nature of the tasks* that are demanded of and performed by the respondents when answering each question and also their reaction to the survey as a whole (Collins, 2003: 230-1). Accordingly, this task-focused model considers four kinds of problems related to the application of a questionnaire: 1) *comprehension problems*, that result from using inadequate vocabulary or complex sentence structures; 2) *validity problems* when respondents interpret the same question in a variety of ways or in the same way as each other but not the way intended by the researcher; 3) *processing difficulties*, resulting from the difficulty in recalling the information requested; and 4) *communication difficulties* between interviewer and interviewee.

Although this cognitive testing approach involves a qualitative analysis, its origins clearly lie in certain types of psychology that employ an underlying cognitive model of response processes, also known as the question-and-answer model, which outlines four actions that respondents have to complete when answering a question – that are related to but not identical with the four problems noted above. Respondents, in this case interviewees, must comprehend the question, retrieve the necessary information from long-term memory, make a judgment about the information needed to answer the question and respond to the question (Tourangeau, Rips & Rasinski, 2000). Although these actions could suggest a linear routing, in practice they involve numerous iterations of and intersections between the different phases, as pointed out by the authors.

2.1. Comprehension

Comprehension covers such elements as how respondents interpret the question, their basic literacy in terms of whether they can understand it or whether there are vocabulary problems that mislead them (e.g. words adults use but children do not). It also includes misinterpreting the response options offered to

them, confusing intended and literal meanings[2], misinterpreting the intended meaning and whether long and overly complex question are difficult to understand. In the case of self-completion questionnaires, it also covers such practical matters as whether the syntax and graphic presentation are clear.

However, reflecting on what is covered by comprehension shows that the term 'cognitive' in 'cognitive testing' is a little misleading, since the word suggests more analytical and mental processing issues, whereas the actual testing covers more than that. For example, besides addressing how interviewees understand the meaning of a question, comprehension also includes the attention they give to the questions and instructions, which could be affected by lack of interest, boredom or distractions during the interviewing. Therefore, the cognitive testing explores whether there are motivated respondents, ones not bored with the interview process, who do not perceive it as being tedious.

2.2 Retrieval

Retrieval refers to the process whereby, having understood the question, the respondent has to recall the relevant information from long-term memory, be it factual or attitudinal. Retrieving factual information – either current or historical - is affected by factors such as the kind of event to be recalled: "the rarer or more distinctive an event is the more likely respondents are to remember. Consequently, commonly occurring types of event will be harder to distinguish and recall individually" (Anderson, 1983, in Collins, 2003: 233). More specifically, similarity is one of the reasons why respondents are not able to recall an event: the item may be difficult to distinguish from similar events or information or the item may be tainted with interference or contaminated from another similar event. Recalling is especially relevant for questions asking about, in this study, technology usage or experiences in the past (as opposed to 'do you have technology X' questions). Whenever respondents search their memory for the relevant information, the cognitive testing looks for difficulties in remembering (which may in turn reflect the fact that the information requested is judged as being not important enough, too distant in the past or too frequent).

2.3. Judgement

Judgement is the process by which respondents formulate their answers to a survey question. This process involves respondents considering whether the question applies to their situation, whether it is asking for information they have, how detailed this information needs to be, how accurate it needs to be, whether they need to modify their answer to meet the perceived needs of the question and so on (Collins, 2003: 233). In fact, these judgments may be made at any stage during the question-and-answer process, they can inform the comprehension, recall and response phases and these evaluations cannot be dissociated

---

[2] For example, one original question in the EU Kids Online survey draft asked if children presented 'another face' when online, meaning did they represent themselves differently. One girl took this literally to mean did she turn her head to a different angle.

from moral considerations, of relevance especially when asking sensitive questions. For example, some respondents may not have thought much about the topic until asked the question. Others may feel ambivalent about what exactly counts as the correct answer, if the options are few but they feel the topic is complex. As already noted, the emotional and moral charge of sensitive topics also affects these judgments and the willingness of interviewees to respond – once again, this means more was being tested than just simply cognitive processing. There is also, for instance, the question of respondents' willingness to perform memory searches if they started to get bored with the interview process or if this is perceived as becoming tedious because it is repetitive (i.e. the respondent already thinks they have already answered a certain question, when in reality they might have been asked a related but different question).

2.4 Response

Response is the final task and involves two components: formatting and editing. Formatting occurs when - having comprehended the question, recalled relevant information and formed a judgment - the respondent has to fit his or her answer into one of the pre-specified answers being offered. Here cognitive testing explores whether there are no appropriate response options to match what the respondents wish to say or whether they have difficulty remembering the response format. In addition to these more practical aspects, editing occurs when respondents want to elaborate their answers before they communicate them or because they may want to conform to notions of social desirability (the interviewee provides what he or she perceives to be the socially correct answer) and are concerned about their self-presentation. Editing can also be influenced by the effects of social distance between the respondent and the interviewer (potentially affecting rapport), concerns about the anonymity and privacy of responses and whether, in the case of sensitive questions, the respondent is comfortable in providing an answer at all.

2.5 Cognitive Interviewing

Among the cognitive testing methods used, and of interest in this article, is 'cognitive interviewing', a tool for pre-testing survey questionnaires that is performed by conducting in-depth, semi-structured interviews with a small number of respondents similar to those targeted in the survey. The interviews are designed to elicit respondents' thought processes when answering the tested question, more specifically, as noted above, how they understood a question and how they arrived at their answer. Current techniques for obtaining this information are the *think aloud* and *probing* approaches: the first makes the respondent articulate their understandings, feelings and judgments, and works better for self-completion questionnaires; the second involves the interviewer asking specific questions designed to elicit how the

respondent went about answering them, and can make the interview easier for the respondent (Collins, 2003: 235).

As the feedback from cognitive interviews is qualitative, analysis of those interviews can indicate the sources of potential response errors as well as various interpretations of the question. However, one limitation already noted in the existing literature (Collins, 2003) is that such interviews can discriminate against less articulate respondents who find it difficult to verbalise their thought process. Because of this, cognitive testing interviews involving children as respondents have to be particularly sensitive to aspects such as their body language.

As Bryman (2004: 159) notes, the desirability of various forms of piloting is not solely to do with trying to ensure that survey questions operate well: piloting has also a role in ensuring that the research instrument as a whole functions well. Of course, cognitive interviewing could be also considered as an example of a 'dialectic method' (Guba, 1990; Lobe et al, 2007), allowing researchers to keep open the communicative channels with their informants. However, the main value of cognitive testing is held to be that by conducting a comparative analysis of cognitive interviews, it is possible to identify not just problematic questions that need to be restructured or more systematic patterns of error, but also patterns of interpretation and willingness to respond by different groups of people. The EU Kids Online II cognitive testing extended the comparison to a cross-cultural level, allowing the researchers involved to identify, even at this stage, some of the commonalities and differences in children's online culture across national contexts and languages.

3. The Role of Cognitive Testing when Designing a Survey Targeted at Children and involving sensitive questions

3.1 The Process of Design in EU Kids Online II

Having set the scene in terms of the significance of cognitive testing in general, this next section provides an overview of the EU Kids Online II project in which the particular cognitive testing reported here took place in order to appreciate the processes by which the whole survey evolved.

The first step was to identify the broad areas we wanted to examine, both in terms of which risk areas we wanted to explore and what types of background information we wished to collect about the child, the parent and the household in general. As regards risk, although there were questions about a variety of areas, the decision was quickly made to focus on cyberbulling, pornography, sexting (sending and receiving sexual messages or images) and meeting strangers offline who had first been meet on the internet (what has been popularly called 'stranger danger'). Hence there were questions about how offline experiences compared to online ones (e.g. bullying versus cyberbullying); whether the experiences were negative (or

not) and if negative and to what degree; and how children tried to cope or deal with the experience. Examples of contextual data to help understand responses to risk would be socio-demographics, psychological profiles of the children, the range of technologies accessed and how they were used, and parental strategies to mediate their children's online experiences. Choosing the particular questionnaires about both risks and contextual data also involved a process of considering what more specific questions had been asked in past questionnaires to explore the scope for eliciting data comparable with previous surveys.

As seen above, topics of the EU Kids Online II survey included questions that addressed private, intimate, upsetting or personally difficult issues related to potential harms. Such 'sensitive' questions have in past research been defined in relation to criteria such as the potential costs of answering particular (sensitive) questions, or in relation to the degree of discrepancy between the responses of people who had to identify themselves and those that could remain anonymous (Crino et al, 1985; Fuller, 1974; Lobe et al, 2008). In the EU Kids Online II study it was felt best to place these particular types of question in a separate self-completion questionnaire. For these, piloting studies as cognitive testing can be particularly crucial since there will not be an interviewer present to clear up any confusion (Bryman, 2006)..

In the next stage of the research the coordinators of the project at the LSE took a lead in producing drafts of the various questionnaire sections (self-completion and administered by the interviewer), but this was very much an iterative process involving the supporting management group and beyond that the wider EU Kids Online II network and international advisory panel. Given the limited time of the interviews, this entailed deciding priorities, working out efficient structures and routing (and avoiding repetition), anticipating problems with answers, etc. At one point the network members even participated, as pairs, in a role play of the interview to discover some of the difficulties that might arise. Throughout there was also interaction with the market research company that would ultimately conduct the actual survey, to draw on their experiences as well as provide fresh eyes to see how drafts from outside the network.

This first stage resulted in a draft face-to-face interview schedule for the child (mainly for the contextual data), a self-completion questionnaire for the child (for the more sensitive areas, including those related to risk) and a parent questionnaire (to gather data about the household, parents' experiences and perspectives and their approaches to mediating their children's online experiences).

After the first draft was prepared there were four more stages, each involving revisions of the questionnaires:

   i.   The first questionnaire (A) went through a process of cognitive testing in the UK. Fourteen children and six parents were interviewed. Four of the children were aged 9-10, four aged 11- 12, five were aged 13-14 and one was aged 15-16. There were eight females and six males. In terms of social

economic status, three parents were ABC1 and there were C2DE (middle- and working-class, respectively). The result of this process was a redesigned questionnaire (B)

ii. Questionnaire B was translated and then subject to further cognitive testing in the remaining 24 countries involving mainly four children per country. Achieving a balance in terms of gender and in living in rural/urban areas were the key factors for children's selection, while this stage gave priority to the youngest children: in 15 countries cognitive testing excluded children older than 10 years old. This choice of age, focusing on those with more difficulty in understanding and answering to the questions (and also less internet experience), might certainly had had impactions for the cognitive testing results. The feedback from this stage led to a redesigned questionnaire C

iii. Finally, there was pilot testing of questionnaire C in five countries (100 people in each of Germany, Ireland, Portugal, Slovenia and the UK), which led to the redesign of what was now the near final questionnaire;

iv. This last questionnaire was then translated (again) but the exercise also raised points that allowed some last, small final adjustments (e.g. question numbering issues).

In the cognitive testing of questionnaires A and B, some questions arose from the national reports reporting feedback from the tests:

1. In the UK, the cognitive testing of English speakers was based on a questionnaire written in their own language: which cognitive problems/difficulties were found?

2. In the remaining 24 countries which cognitive problems/difficulties were found in the translated questionnaires,? Did they coincide with the UK results? Were they different? What kinds of problems, if any, were introduced by the translation from English? Were there problems related to other cultural contexts?

The next section presents the first results from the cognitive testing in the UK and then the results for the other 24 countries, indicating what kinds of solutions were found for the difficulties that were identified

3.2 Cognitive Testing in the UK

The report from the market research agency illustrates the type of evidence that cognitive testing use to explore the issue of children's general reaction to the questionnaire:

"*Children across all ages were happy to engage in the topic and showed very few signs of feeling uncomfortable either when answering questions about their own internet experiences or about parental mediation. Respondents were happy to provide detail of behaviour that might be frowned upon (for example having a social networking profile below the age of 13 or knowing how to get round of parental mediation) and at no point did any respondent ask to skip a*

question or stop the interview prematurely. Children of all ages appeared genuinely interested in the topic and were often keen to provide further contextual information to their responses - this is shown in the fact that interviews were frequently able to go on for longer than 45 minutes. Children were able to get through the long batteries of similarly formulated questions with relative ease, although some commented that they would have been likely to lose interest if those sections had been any longer."

As regards the other elements being tested, as noted above in the outline of cognitive testing principle, sometimes it was unclear whether one should consider the points arising to be comprehension, recalling, judgment and response issues, since they were interconnected. Moreover, it was not just the case of children misinterpreting questions, there were also real ambiguities in the drafts as revealed when the children thought of scenarios that then required the design team to be more precise in formulating the questions.

For example, one point related to the time frames of particular questions. When originally asked '*do you use the internet...*' in a variety of locations one first response from several children was that they used to use it in certain places (e.g. cybercafé) but had not done so for some time. Hence, '*these days*' was added to indicate an interest in their current practices. But apart from the need to be more specific, when asked how often they used the internet and then probed some respondents thought '*a week*' referred to the Monday-Friday school/work week. They chose an option '*more than once a week*' but when it was pointed out that a week could include a weekend, changed their choice to '*most days*'. This was an example of a general challenge thrown up by some of the testing since it would have made the questionnaire too long if we were to continuously add clarifications such as 'what counts as a week', and we had to simply be aware that some misinterpretation was possible.

Definitions were also an issue when children asked 'what does x include'. For example, when asked simply about their use of '*the internet*', many children did not interpret this as had been expected: initially they did not think about emailing and instant messaging, two common internet practices. Hence it was necessary to add a prior clarification '*By using the internet I include....*'. But related questions about what was being covered in a question emerged throughout, as when some children asked '*Does downloading software include "downloading games?"' 'Is MSN classed as a chatroom?*' Providing nationally relevant examples (throughout) helped to steer the children in right direction (in the case of UK social networking sites, *Facebook, Twitter,* and *MySpace*). Another solution was to add definitions to the interview schedule, but definitions that the children would not actually see unless they specifically requested clarification. For example, the instructions for the interviewer were '*If child asks: by social network I mean a site where you have a profile and can meet other people or post a message on your or their profile*'. There were, as noted

above, limits to how much clarification one can give, here shown by the fact that some children did not understand words like 'profile' or 'personality'. Therefore, in the questionnaire in general the instruction was added '*If you don't know what something is, don't worry about it – just say you don't know*'. Lastly, definitions did not only relate to parts of the internet – when asked about use when 'adults' were in the room, one child asked about the status of his 17 year old cousin – hence the question had to clarify that this meant 18 years or over.

Ownership presented some challenges, especially because the literal ownership (e.g. who bought a technology) does not correspond to the realities of usage. As one child pointed out, although he had not bought the laptop and hence it was classed as his parents', he was pretty much the only person who used it in practice and therefore thought of it as his 'own'. Another chose the option indicating she used '*own laptop*' rather than '*shared laptop*,' even though it was her sister's because someone did indeed 'own' it. The clarification here involved changing '*own*' to '*your own*'. A similar point emerged when asking if children had their '*own profile*' on a social networking site. For example, one child shared a profile with her old cousin and another used her mother's account to speak to family and friends. It proved difficult to respond to ever single scenario, such as these. What the cognitive testing revealed in this case was that in interpreting the findings, analysis would have to be aware the statistics might miss such proxy use given this question wording.

Counting or calculating (e.g. frequencies) is another example of how recalling is interlinked to other cognitive acts. For example, when asked how frequently they used the internet (the questionnaire already made a distinction between school and non-school days) the children understood and could manage to choose from response options, but it took them some time to work it out. When we later had to reduce the length of the interview this question was important enough to keep (although on the request of some interviewees '*on average*' was changed to '*typically*'). Nevertheless, this type of information from cognitive testing (i.e. the time taken to answer) was a consideration when choosing which questions to prioritise. When asked how many contacts they had on social networking sites, those conducting the cognitive testing observed that boys founds this easier and could answer more quickly, one boy pointing out that this was because he and his male peers were competing to see who had the most contacts. By comparison, half the girls found this more difficult, sometimes observing that this was something they did not keep track of. Again, the cognitive testing provided useful background knowledge for interpreting the results, not only for designing the questionnaire.

The questions on rules in the parental mediation section revealed an unanticipated set of social practices that required a reformulation of the questions asked. The original question was '*Which of these things, if any, are you forbidden by you parents to do on the internet?*' (followed by a list). Cognitive testing showed

that there were often cases where the picture was not so black and white, and that children could do some things if they were given permission or if the parents were sitting with them when they went online. This meant changing the response options to '*Can do anytime*', '*Can only do with permission or supervision*' and '*Can never do*' (as well as '*Don't know*') in order to explore the nuances of rules. That same testing also showed the need to reword '*forbidden*' to '*let you do*' because both parents and children had a very strict interpretation of the word 'forbidden': i.e. that something was not allowed at all, and parents had explicitly stated a rule about this. But in practice a number of parents noted that even though there were rules that they hoped the children understood, they had never used the word 'forbidden', nor had they always explicitly (and categorically) stated that there were certain rules - instead they hoped that the implications of what they had said at some stage would be absorbed by the child. This underlined a lesson from several of the examples. Up to a point, one can try to search for better words to use in formulating a question. But cognitive testing also revealed the areas such as this where there is more scope for ambiguity, whatever the statistics show - in this example when deciding if rules exist or not. Hence the need to add caveats about this and certain other areas, when reporting results.

As regards general familiarity with internet terms, the cognitive testing in the UK produced a similar result to the findings of the second wave of testing in other countries. Younger children especially (and some older ones) tended not to recognise generic descriptions such as '*instant messaging*', '*social networking*' or '*virtual worlds*'. Instead children often referred to specific brand names such as *MSN*, *Facebook* or *MySpace* when talking about online different environments and practices. Hence in the interviews nationally relevant brands were consistently added as examples.

Some words were very strategic in the survey and so were of particular interest for cognitive testing, including checking whether they were understood in the same way in different languages. In particular for the risk section, it was important to find a form of words that indicated at least a slightly negative experience because these would be used in various filters – if children said *no*, (the experience was not negative) at certain points in the self-completion questionnaire they then skipped further questions on that particular risk; if they said '*yes'* (it was negative), the subsequent questions explored the nature and extent of that negative experience. Therefore the initial terms had to be reasonably mild to set a low threshold in order to route children to the follow up questions. In the end, the key term was whether the children were at all '*bothered*' by an experience, together with related words like '*upset*', '*worried*' or '*uncomfortable*', each with slightly different connotations but all indicating this low threshold. Although the terms were understood, they could mean slightly different things for different children. One Male, aged 10, noted '*Upset or worrying is something that is threatening where you don't really like it and it is not normal.*" Meanwhile a male, aged 15 commented "*It is something that you haven't prepared yourself to*

see. You expect to see porn everyday, nothing will upset or bother me anymore." Ultimately the term '*bother*' was so vital that a number of lines of clarification were added to indicate what it did or did not imply (e.g. '*it does not just mean "bad news" as when a favorite football team has lost a game*').

Let us now looking at the translated questionnaires and the cognitive testing results in the remained 24 national reports.

3.3 Cognitive Testing in 24 Countries

As noted earlier, cognitive testing of the next draft of the questionnaire (version B, incorporating contributions from the UK testing and translated into the various national languages) was next conducted in all the remaining 24 countries. Five main difficulties emerged from the national reports from this stage: the length of the questionnaire; unfamiliarity with certain words, especially internet terms and uses; embarrassment when faced with sensitive questions; and the translations of the strategic filter questions (e.g. upset); and routing.

Based on the previous UK testing, the ambition of providing as many as possible unambiguous and repeated definitions of what was asked, had resulted in a questionnaire that the children felt was simply too long, with excessive and repetitive instructions to be read and understood by children, some of who had relatively low reading skills. This all contributed to the outcome that many children across the countries become frustrated after a certain time point in the interview. Facing this difficulty, in the final questionnaire the amount of text in the introduction was reduced as much as possible and repetitions were avoided – as will be clarified in more detail in the next section on key lessons learnt from this exercise.

The second difficulty related to many children's unfamiliarity with the English internet terms that are currently incorporated and used in other languages, such as *download*, *pop-up, to post, software, chat-room, bookmark, cookies* or *blogs.* These terms appeared in a section of the questionnaire aimed at identifying advanced internet competencies and practices. Although the questionnaires had incorporated short descriptions of their meanings, provided by the national teams, the questions were still not understood well enough by the youngest children. As various national reports from the cognitive testing also pointed to the fact that the younger children probably did not yet have the competencies being tested, the solution for the final questionnaire was to miss out this section when interviewing children aged 9-10 years old.

But English internet terms were not the only problem. For example, the English expression '*X-rated*', associated with the classification of media contents, proved to be unfamiliar to many. In the self-completion questionnaires, '*adult/X-rated websites*' indicated one of the places where children could find sexual images. Among the few countries where questions on sexual risks were answered, none or almost none of

the children knew its meaning (or even the meaning of the term '*adult website*'). In the two countries where the expression was known by children, their reactions seem have been affected by the different connotations and uses of the word in the respective countries: The Hungarian report noted *"No embarrassment. X-rated is known and understood; it is a known expression from TV. However, the content of these web pages and films are not really known".* Meanwhile the Dutch report noted "*X-rated" is understood but children are a bit embarrassed when they hear it."*

Another example related to a question about whether children had visited websites discussing bulimia and anorexia. In the UK testing, while a few children were not sure of the exact meaning, most recognised that these terms related to eating disorders. In the 24 country testing, the terms were far less familiar to many children.

The third shared problem in these cognitive testing was young children's embarrassment when asked sensitive questions, specifically in certain sexual risks sections of the self-completion questionnaire. While reports about the older respondents pointed to some difficulties with internet terms and time scales, reports on the younger ones often drew attention to their discomfort, distress and unwillingness to answer, expressed mainly by body language. This attitude was cross-cultural to the extent that it was noted in Northern and Southern European countries. The solution to this problem was the division of the self-completion questionnaires by age, one for aged 9-10, and other for the aged 11-16. The first one, for the youngest children, dropped the sexting section.

As noted above, in the UK sensitive terms such as *bother*, *upset* or *worry* were well understood, although they could mean slightly different things for different children. The fourth problem identified was that finding the corresponding terms proved to be a difficult translation task in almost all languages, due either to the absence of distinctions between different terms in some languages or their additional connotations. As the Swedish cognitive testing report noted *"They were not sure on the meaning of upset (upprörande) they interpreted it differently. They do have a different perception of the meaning of this one. The word can mean different things for different persons – it is a quite broad word. Not sure how to adjust it".*

Opposing considerations as regards trying to find the adequate translation of the word '*worry*' emerged even among very close languages (here, different versions of French). The Belgium report pointed out that *"The translation of 'worrying' (in Belgium) should be 'dérangé' and not "ennuyé (as in France)."* Meanwhile the French report noted *"Worrying and upsetting meant the same thing for all the children."*

Other reports suggested that children's judgments - emotionally charged in response to the connotations of sensitive words - might have affected their responses to certain questions. When asked about negative internet experiences in general ('*In the PAST YEAR, have you seen something on the internet that you found upsetting or worrying?*') an unwillingness to reveal certain things about themselves may well have

produced some negative answers, which should therefore be considered when analysing the final results. For instance the Spanish report on four respondents aged 9-10 noted that "*Both 'upsetting' and 'worrying' are understood perfectly. However the question might sound too 'strong'/ dramatic and in both cases the child said 'no'.*"

The Dutch report observed "*The children interpret 'upsetting' to mean: something that is not nice/not fun; when you are scared when something suddenly appears on your screen, something 'dirty'; something scary that you are not supposed to see (monsters). When asked what upsetting is they can give an explanation. However, most children filled out 'don't know' or 'don't want to say' instead of 'no'. So they are embarrassed and don't want to tell.*"

Finally, routing was the other major issue in the risk self-completion section, identified originally in the UK testing and, despite amendments to the questionnaire, in the subsequent wave of cognitive testing. In most countries the survey was conducted on paper, and at various filter questions the child would sometimes jump ahead prematurely to the next risk sections. Despite the best efforts to be clear in the first draft, the cognitive testing revealed significant routing problems in all countries: some children chose not to read the introductory explanations or routing instructions, others were confused about which question they should go to next. Also due to the amount of time spent on the self-completion part, some children lost interest and motivation, as noted in these reports. For instance the Portuguese report on four 9-10 years old respondents observed "*More or less at this time they were very tired, and asked if they would be missing too much if the finished at that point. They understand the routing but they always confirm with us.*" Meanwhile the Romanian report noted "*One of the children was not able to follow any instructions or filters on his own – he needed to be permanently assisted by the interviewer.*"

Ultimately this difficulty required multiple solutions in the final questionnaire. Routing was reduced, meaning some loss of information that would have been of interest. There was also a compulsory practice question with the interviewer at hand, to help children understand the routing rules. Younger children (9-10) who had fewer questions anyway, received booklets with some routing clues in red (which was more costly to produce). The risk sections were labeled and children instructed to answer each section. And various other clues, reminders and details of the format were altered in order to try to minimize these routing problems.

3.4 The Lessons from Two Sets of Cognitive Testing

This exercise of conducting cognitive testing twice was particularly revealing, even if it delayed the whole timetable and created substantially more work for the questionnaire designers. In short, the questionnaire redesign after the first test itself created new problems that were picked up in the second test. After the

first, UK, test a number of respondents interview asked for more material in various sense. In the case of questions offering scales (e.g. *once a day, once a week, once a month*), some children requested finer gradations to more accurately reflect their situation. If the options were not on a scale, some children simply wanted more options to reflect their circumstances, rather than having just a category 'other'. Some asked for more clarifications or definitions to accompany the question. And in some case it seemed useful to repeat a statement because the respondents might have forgotten it still applied (e.g. '*by the internet we mean…*'). All these were added, and it became clear at the second stage of cognitive testing that these additions themselves contributed to making the questionnaire too long!

There is a practical issue here: the project paid the market research agency for a certain length of questionnaire, measured in terms of time taken to complete it (in this case, 30 minutes on average across countries for the child). The fact that the questionnaire was now too long was revealed at the later piloting stage, one of whose functions was to measure the interview timing. But in addition, and as noted above, the second phase of cognitive testing had by then already indicated that a number of children, particularly but not only younger children, were getting bored. In fact, some were verbally complaining about the length of the questionnaire and, (since some statements were now repeated on purpose), its repetitiveness in places.

This feedback from the second cognitive testing and piloting ultimately led to a number of changes. In some cases the options were reduced back to the original number or otherwise shortened – some people had to simply put their answer under the heading '*other*', or choose the nearest option on a scale even if it was not ideal. Sometimes younger children had even fewer options (e.g. '*yes*' or '*no*', rather than the richer information we might desire from using a scale). A number of the repeat statements that had been added were removed. While some clarifications were left in the text, others were moved to the interviewer's instruction sheet so that he/she could supply explain a point further if requested - which meant that the majority of children did not have to have this read out if the meaning was already clear to them. And in some cases, as noted earlier, questions for the 9-10 year olds were dropped and only asked of 11-16 year olds. Admittedly there were often other considerations at work in this choice, such as parents thinking these questions were too sensitive for 9-10 year olds (e.g. concerning sexual matters). But the aim was also to reduce the length, especially in the self-completion section, for younger children who appeared to be slower readers and more likely to lose attention quickly. In sum, repeating the cognitive testing reminded us that while all the points raised the first time were valid and addressed in good faith, there are always trade-offs and the second round showed how the corrections then needed themselves to be amended or in some cases dropped. One key lesson, especially for future studies that only have one round

of cognitive testing, is that the feedback may be useful but one has to think carefully about the implications of responding to all the issues raised.

3.5 The Limits of Cognitive Testing in the EU Kids Online Study

Cognitive testing in general is conducted with a relatively small number of people, since they are being asked quite detailed questions about how they are responding to the questionnaire. In fact, in the EU Kids Online II testing no one child went through all the questions since that would have taken far too long – different children answered and commented on different sections. The problem for the project was that in many countries the sections on risks could not be tested because the filters routed out the children involved in the test. It was clear in advance that for some areas of risk only small proportion of children will have encountered them (e.g. meeting strangers online) and hence go on to answer the follow up questions about the experience. In a sample of 1000 it was anticipated that this would pick up enough children for further statistical analysis (more so in some areas, like cyberbullying or encountering pornography). But in a sample of just a few, only some of whom were answering this section, in many countries there were simply no children involved in the testing who had had (or were willing to admit to) experiences that would have led them to the follow up questions. This also related to the fact that we particularly tested the material with 9-10 year olds since we anticipated that they would have the most comprehension problems or get bored quickly. And these are the ones who are probably less likely, compared to older children, to have had some of these risk experiences.

4. Conclusions

Are there alternatives to cognitive testing? While piloting a survey before conducting the fieldwork may raise some of the issues elicited here by cognitive testing, piloting itself is in general not designed to explore responses in such depth. Indeed, piloting usually tries to replicate the experience of the survey with little time for probing further, since one aim is often to measure how long the interview takes. Another practice sometimes followed in research is to try out informally a draft questionnaire, on friends or family members for example. That too might elicit some of the points raised here, but the formal cognitive testing is likely to be far more systematic in terms of specifying in advance all the questions asked about the experience of answering the questions. This is, of course, necessary for a survey to be tested in a comparable way across countries.

In sum, there are sufficient illustrations from this project to demonstrate the value of taking this extra step of formal cognitive testing in the research process, to check both understandings of questionnaire and

reactions during and to the interview process. Moreover, while quantitative research sometimes builds on previous qualitative studies where this is not the case, to a large extent true for some of the EU Kids Online II areas, cognitive testing can also reveal some underlying social realities (for example about rules) that can lead to a reformulation of the question and answer options. There are, of course costs involved in cognitive testing: both financial and in terms of the extra time it takes. But on reflection the experience of EU Kids Online II study would find this justified - it is clearly a very worthwhile stage to add to the design process.

However, the experience of trying out cognitive testing, shows that here, as in the rest of the questionnaire design process, there are decisions and trade-offs that require careful thought. That was clear in the decision to focus the second wave of tests on younger children (because of comprehension considerations) that may well have reduced the number who could test certain risk questions. It was also clear when the second wave of test showed how responding to the first test had itself caused different problems.

Lastly, and returning to the theme of the introduction, cognitive testing can be considered to be an example of combining using a qualitative approach for 'instrument development' in a quantitative study – i.e. it can produce 'better' questionnaires and interview processes, at various levels and with various solutions. However, the EU Kids Online II case study shows how in the process of doing so cognitive testing also reveals the limitations of those same questions and informs the analysis beyond the design stage – despite the fact that the literature reviewed focused mainly on design. Sometimes one cannot cater to every scenario raised by participants in the testing. Or else it becomes clear that even with the best form of wording in the circumstances (given the limitations of the length of the interview) the questionnaire may fail to fully capture certain aspects of social life or produce certain social responses. Hence beyond 'instrument design', cognitive testing alerts researchers to these situations, to possible social processes at work in the interview, to the complexity of what is being studied, which can then all be taken into account after the design and fieldwork, when in the stage of analysis.

References

Bryman, Alan (2004). *Social research methods.* Oxford University Press, 2nd edition

Bryman, Alan (2006a) Paradigm peace and the implications for quality, *International Journal of Social Research Methodology*, 9 (2), 111-126.

Bryman, Alan (2006b) Barriers in integrating qualitative and quantitative research, *Journal of Mixed Methods Research,* 1, (1), 8-22.

Bryman, Alan. (2006c) Integrating quantitative and qualitative: how is it done? *Qualitative Research 6* (1), 97-113.

Collins, Debbie (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12,* 229-238.

Crino, M.D., Rubenfeld, S.A., Willoughby, F.W., (1985). The random response technique as an indicator of questionnaire item social desirability/personal sensitivity. *Educational and Psychological Measurement*, *45*, 453-468.

Fuller, Carol (1974). Effect of anonymity on return rate and response bias in a mail survey. *Journal of Applied Psychology, 59*, 292-296.

Guba, Egon (1990). The alternative paradigm dialog. In E. G. Guba (Ed.), The Paradigm dialog. Newbury Park, Calif; London: Sage Publications.

Green, J.C., Caracelli, V.J., and Graham, W.F (1989) Towards a conceptual framework for mixed-method evaluation design. Educational Evaluation and Policy Analysis, 11 (3), 255-74.

Jensen, Klaus BBruhn Ed. (2002). *A Handbook of Media an Communication Research. Qualitative and Quantitative Methodologies*. London, Routledge.

Lobe, Bojana., Sonia Livingstone, et al. (2007). *Researching children's experiences on the Internet across Countries: Issues and Problems in Methodolog*y EU Kids Online Deliverable D4.1 Methodological Issues Review. EC Safer Internet Plus Programme

O'Sullivan, Tim, John Hartley, et al. (1995). *Key Concepts in Communication and Cultural Studies*. London, Routledge.

Tourangeau, Roger, Rasinski, Kenneth. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.