eukidsonline.net

EU Kids Online 2020: Technical report

Rostislav Zlamal, Hana Machackova, David Smahel, Katarzyna Abramczuk, Kjartan Ólafsson and Elisabeth Staksrud

Please cite this report as:

Zlamal, R., Machackova, H., Smahel, D., Abramczuk, K., Ólafsson, K., & Staksrud, E. (2020). *EU Kids Online 2020: Technical report*. EU Kids Online. https://doi.org/10.21953/lse.04dr94matpy7

Authors:

Rostislav Zlamal, Interdisciplinary Research Team on Internet and Society, Masaryk University

Hana Machackova, Interdisciplinary Research Team on Internet and Society, Masaryk University

David Smahel, Interdisciplinary Research Team on Internet and Society, Masaryk University

Katarzyna Abramczuk, National Information Processing Institute, Poland

Kjartan Ólafsson, Faculty of social sciences, University of Akureyri; Department of Media and Communication, University of Oslo

Elisabeth Staksrud, Department of Media and Communication, University of Oslo

Authors' acknowledgements:

This report and the development of a joint comparative dataset was partially supported by the project FUTURE (GX19-27828X) which is financed by the Czech Science Foundation, and a grant from the Norwegian Ministry of Justice and Public Security's Proposition 12 S (2016–2017) Escalation Plan against Violence and Abuse (2017–2021). The authors acknowledge the support of all members of the EU Kids Online network. We thank Eliska Dufkova for her help with the finalization of this report and for her thorough work on the development of the Data Dictionary.

Full findings from the EU Kids Online survey are described in Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries.* EU Kids Online. https://doi.org/10.21953/lse.47fdeqj01ofo.

The EU Kids Online network is a multinational research network. It seeks to enhance knowledge of European children's online opportunities, risks and safety. It uses multiple methods to map children's and parents' experiences of the internet, in dialogue with national and European policy stakeholders. Now working in more than 30 countries, the network integrates research expertise across multiple disciplines and methods.

The project EU Kids Online maps the internet access, online practices, skills, online risks and opportunities for children in Europe. Teams of the EU Kids Online network collaborated between autumn 2017 and summer 2019 to conduct a major survey of 25,101 children in 19 European countries.

For all reports, findings and the technical report of this survey, as well as full details of national partners, please visit **www.eukidsonline.net**

Contents

Acknowledgements

EU Kids Online members acknowledge the support of the following institutions which enabled the survey to be carried out in 19 countries. We also thank all the research agencies, teachers and school staff who helped to administer the survey and helped in the research process, as well as all the children and young people who participated.

Belgium – Flanders: University of Leuven, Institute for Media Studies.

Croatia: Association for Communication and Media Culture, Agency for Electronic Media of the Republic of Croatia, Croatian Telecom Inc., City of Zagreb, Croatian Regulatory Authority for Network Industries, Center for Missing and Exploited Children and Ipsos Puls.

Czech Republic: Faculty of Social Studies, Masaryk University.

Estonia: For generous financial support: Ministry of Social Affairs, Ministry of Education and Research (from the European Social Fund), Ministry of Justice, Estonian Internet Foundation, and Institute of Social Studies, University of Tartu (research projects PUT 44, financed by the Estonian Research Council, and IUT 20-38, financed by the Ministry of Education and Research); for data collection: Turu-uuringute AS.

Finland: Funded by the Finnish National Audiovisual Institute during 2017-19.

France: OpinionWay with funding from Facebook and support from the International Observatory of Violence in Schools and the Université Nice Sophia Antipolis.

Germany: UNICEF, Deutsche Telekom Stiftung, Medienpädagogischer Forschungsverbund Südwest and Niedersächsische Landesmedienanstalt (NLM).

Italy: DG for the Student, Participation and Integration of the Ministry of Education, University and Research (MIUR) and realised by OssCom- Research Centre on Media and Communication.

Lithuania: Children's and adolescents' Internet use in Lithuania: possibilities and risks tendencies in EU context (No. S-MIP-17-1/LSS-250000-1087, Research Council of Lithuania), Institute of Psychology at Vilnius University.

Malta: Malta Communications Authority, Directorate for Learning and Assessment Programmes, Besmartonline!, Tech.mt, Commissioner for Children, Foundation for Social Welfare Services, Cybercrime Unit.

Norway: National State Budget 2017-18 under the Ministry of Justice and Public Security's Proposition 1S (2016-17) and Proposition 12 S (2016-17) Escalation Plan against Violence and Abuse (2017-21).

Poland: Orange Foundation.

Portugal: DNS.PT Association, for funding the field work, and DGE - Directorate-General for Education, for support and communication with schools.

Romania: IRES (Institutul Român pentru Evaluare şi Strategie), Institute of Sociology and the Digital Lives. Research, Education and Intervention platform.

Russia: Russian Science Foundation (Project no. 18-18-00365).

Serbia: Institute of Psychology, University of Belgrade, Department of Media and Communication, University of Oslo, Norway, Ministry of Education, Science and Technological Development in Serbia, Ministry of Trade, Tourism and Telecommunications in Serbia, OSCE Serbia, UNICEF Serbia, and USAID Serbia.

Slovakia: The Ministry of Labour, Social Affairs and Family of the Slovak Republic within Operational Programme Human Resources Development 2014-20 'Národný projekt Podpora ochrany detí pred násilím' (ITMS 2014+: 312041M679).

Spain: Spanish Ministry of Economy and Innovation (MINECO) through the project CSO2017- 88431-R. The editing of the Spanish report was cofunded by the Spanish Cibersecurity Institute (INCIBE) and SIC-SPAIN 'Safer Internet Centre Spain' funded by the EC (CEF-TC-2018-1).

Switzerland: Bundesamt für Sozialversicherungen – Nationale Plattform Jugend und Medien and Pädagogische Hochschule Schwyz – Institut für Medien und Schule.

About this report

This report describes in detail the methodology used for the EU Kids Online IV project (see the description of the four phases of the project in the next section). Within this project, a large-scale survey of children aged 9–17 from 19 European countries was conducted. The data were collected between autumn 2017 and summer 2019 from 25,101 children by national teams from the EU Kids Online network.

This report provides information about the nature of the project, how the questionnaire was developed, sampling and data collection, ethical issues, data management and weighting. The information in this report should enable dataset users to understand the logic and nature of the survey.

For dataset users, we also recommend using the 'Data Dictionary' (available at eukidsonline.net), a related document that systematically maps all the information related to the data in the dataset. Moreover, Annex 2 of this report provides concise key guidelines for dataset users. We highly recommend using these short guidelines during work with the EU Kids Online 2020 dataset. Annex 3 contains a description of the key variables. Full questionnaires and their national forms are available at eukidsonline.net.

EU Kids Online project

EU Kids Online is an international interdisciplinary research network that seeks to enhance knowledge about European children's online opportunities, risks and safety. The network integrates research expertise across multiple disciplines and methods in more than 30 countries. It sets out to provide empirical evidence on children's, young people's and parents' online experiences.

So far, EU Kids Online has been comprised of four waves. Between 2006 and 2009, the *EU Kids Online I* initiative identified and critically evaluated the findings of nearly 400 research studies, drawing substantive, methodological and policy-relevant conclusions. Between 2009 and 2011, *EU Kids Online II* conducted a representative survey across 25 member countries with national samples of children aged 9–16 and their parents. The aim was to produce a rigorous, cross-nationally comparative quantitative evidence base regarding internet use across Europe. This phase of the project was undertaken by the EU Kids Online network, comprising more than 70 experts focused on the social uses of the internet and new media; media education and digital literacy; childhood and family studies; the psychology of adolescence and identity; legal and regulatory perspectives; and research methods. From 2011 to 2014, in *EU Kids Online III*, qualitative investigations were conducted in nine countries to provide an in-depth and contextualized understanding of the quantitative findings. In the fourth wave, *EU Kids Online IV*, from 2017 to 2019, the network designed a second representative survey of children and online risks and opportunities. The survey was conducted in 19 European countries and targeted children aged 9–17 who use the internet. This report describes the methodology related to the fourth wave, that is, the EU Kids Online IV survey.

EU Kids Online IV: aims and principles

In line with the overall project, the fourth wave aimed to provide an understanding of the online activities and risks experienced by children, with a specific focus on those aged 9–17. A theoretical framework for research on children's online experiences was revised and enhanced. The network has also continued to update the EU Kids Online public database, documenting and coding recent and updated evidence about children's use of new media across Europe. Furthermore, EU Kids Online members have initiated new collaborative cross-national projects on special topics (e.g., young children and online use, cyberbullying etc.). Findings are published in EU Kids Online short reports and disseminated within national, European and international research forums, and among national, European and international stakeholders.

The core part of the fourth wave was an international survey, which differed slightly from the survey carried out in the second wave. Specifically, this survey was not directly centrally coordinated, and national teams organized funding and data collection at national level. However, several principles were established to ensure unified approaches that would maximize the comparability of national surveys. A general methodological approach and specific guidelines were formulated. These included the sampling strategy, the form of the questionnaire, translation procedures, data management, data analysis and the reporting of the findings. Individual national teams were provided with these guidelines and a unified matrix for data entry. Communication links between the national teams and the EU Kids Online Management Group were established. In order to ensure that we obtained a solid base for international comparisons, the national sampling procedures and questionnaires had to be approved by the EU Kids Online Management Group. Any country-specific challenges during the survey, data cleaning or merging could be consulted and resolved individually.

EU Kids Online IV questionnaire

This section describes the development and nature of the measurement tool used in the EU Kids Online IV survey. It includes a description of the process of developing the questionnaire as well as its structure (the full questionnaire in English as well as its translation into other languages is available at www.eukidsonline.net).

The development of the new questionnaire was based on the joint work and expertise of members of the EU Kids Online network, led by Professor Elisabeth Staksrud (University of Oslo, Norway) and researcher Kjartan Ólafsson (University of Akureyri, Iceland). Most items (especially in the core and extended core sections) were based on the questionnaires from the EU Kids Online II survey (see www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online/toolkit) and the Global Kids Online survey (see http://globalkidsonline.net/tools/). The questionnaire for EU Kids Online IV was designed to reach maximum comparability with both surveys. Nevertheless, the questions were updated according to the current state of technology and internet usage. Members of the EU Kids Online and Global Kids Online networks discussed the shape of the questionnaire. The final version was approved by the EU Kids Online Management Group in September 2017. The translation of the questionnaire was coordinated and supervised by expert members of the EU Kids Online network within each country. In several countries (Czech Republic, Estonia, Italy, Lithuania and Switzerland) cognitive testing was conducted to assure the comprehensibility of the questionnaire and its national translation.

Development of the questionnaire and its basic structure

The EU Kids Online IV questionnaire was divided into several sections, and in the data, a fourth type of question was specified, the country-specific questions.

Core questions, extended core questions, optional questions and country-specific questions

- Core questions were mandatory for all countries that aimed to be part of the international dataset. They were intended to be used for central cross-country comparison. However, some were omitted in several countries (for details see the 'Data Dictionary'). Questions were designed to correspond with the list of core questions in EU Kids Online II and Global Kids Online surveys.
- Extended core questions were extended to several topics within the core questions and were not mandatory.
- Optional questions were further extended to selected topics or covered other research interest areas and were not mandatory.
- Country-specific questions were slightly modified by a specific country, e.g., by adding another value to the question or slightly changing the meaning after translating it into the country's language (for more details about individual modifications, see the 'Data Dictionary').
- The countries were instructed to use all of the core questions and to choose from the optional questions in line with their preferences.

Non-mandatory modules

Several topical modules were developed independently from the core questions in order to capture current themes in society and policy-making. Each country decided whether or not to include the module(s) in their national survey. The modules are:

- Module M1: 'Cyberhate', responsible person: Catherine Blaya
- Module M2: 'Bystanders of cyberbullying', responsible person: Hana Machackova
- Module M3: 'Digital citizenship', responsible person: Tijana Milosevics
- Module M4: 'eHealth', responsible person: David Smahel
- Module M5: 'Internet of things', responsible person: Giovanna Mascheroni

Table 1 shows the different modules used in participating countries. Several countries didn't use all of the questions from the individual modules (for the number of items used, see the 'Data Dictionary'). No country used Module M4.

Table 1: Optional modules and interview length

Country	Module	Expected time of interview (in minutes)
CH	None	–
CZ	M1	50
DE	M1, M2	45
EE	Part of M3	60
ES	None	–
FI	M1, M3	N/A
FR	M1	40
HR	None	–
IT	M1, part of M2	55
LT	Parts of: M1, M2, M3	65
MT	None	–
NO	Parts of: M1, M2, M3, M5	60
PL	Part of M1, M2	45
PT	M3, M5	40
RO	M1, M2	35
RS	None	–
RU	M5	50
SK	M1, M2	45
VL	M1	50

Optional questions for younger children

To account for the complexity or sensitivity of some of the questions and the overall length of the questionnaire, selected items were proposed as optional for younger children. Each country decided which questions should not be asked of younger children. In most countries (except Spain, Finland, Croatia, France and Flanders) the questionnaire was distributed in two forms: a full version for older children and a shorter version for younger children. The category of 'younger children' consisted of 9- and 10-year-olds. In some countries, however, the definition of 'younger children' differed from the recommended one (i.e., 9–10). Norway used a different age range (see Annex 1); in Malta, Lithuania and Portugal, the younger age group was defined as 9- to 11-year-olds; and Germany and Estonia defined multiple age groups for administering different types of questions . Details can be found in the 'Data Dictionary' and Annex 1.

System of coding for the questionnaire and variables

This section describes the logic behind the coded names of the questions in the questionnaire, the dataset and the 'Data Dictionary'.

Topical areas

The survey consisted of several topical areas identified by a letter code in their variable name (e.g., in the form 'QX' where 'Q' stands for 'question' and 'X' stands for survey part code). These identification codes were used in the dataset and in the 'Data Dictionary'. They include the following areas:

- **Child identity and resources** (A)
- **Access and use** (B)
- **Opportunity and practices** (C)
- **Digital ecology** (D)
- **Skills** (E)
- **Risks** (F)
- **Well-being** (H)
- **Family** (I)
- **School** (J)
- **Peers and community** (K)
- **Modules** (M)

Prefixes

Prefixes were used to differentiate the types of questions, which is especially helpful in differentiating the core questions that were intended to be used in all the countries from the optional questions. Prefixes used include:

- **c_** – core questions
- **ec_** – extended core questions
- **op_** – optional questions
- **m1_**, **m2_**, **m3_**, **m5_** – questions of modules
- **NO_**, **SK_**, **DE_**, **FR_**, **LT_**, **PL_**, **VL_** – the prefix consisting of country identification labels a country-specific question in the dataset and the 'Data Dictionary'. Specifically, Norway, Germany, France, Lithuania, Poland, Slovakia and Flanders made some changes to some of the questions (mostly by adjusting the answers) (see the 'Data Dictionary'). In the dataset, a missing value in the original question -91 specifies that the respective question was asked in the modified version and the data was not included in the original variable (see section 'Missing values').

Suffixes

- **_rt** – this is used to distinguish routed items. 'Routing' means that selected follow-up questions were asked only if previous answers met specified conditions. This was most commonly used for 'risks' questions, where more details were asked only of children who answered that they had experienced the risk in question. The values of routed-out questions were coded as -96 (see section 'Missing values'). More layers of routing were marked with numerical values depicting the respective layers (_rt1, _rt2…).
- **_oy** – used for questions that were optional for use with younger children (9- to 10-year-olds) due to complexity or sensitivity. If _oy questions were not asked in a respective country, they were coded as -93 (or -92; see section 'Missing values').
- **_rec** – labels derived variables, that is variables created out of the original item; in most cases, the variable included data from those that were routed out and gave them meaningful value within an originally routed-out variable (e.g., those children who said 'No' to the original question were given the value of 'Never' in the derived one).

Missing values

Several types of missing values were used to identify missing data in the dataset. They comprise the missing values specified in the questionnaire, in which each question included the options 'I don't know' and 'Prefer not to say', and missing data from data collection and data management procedures. Each national team was instructed to follow the differentiations of missing values during national data collection and national data management. These were checked and verified during centralized data-cleaning procedures. During this procedure, additional missing values (-95 to -91) were defined and coded to specify data missing due to technical and structural errors.

- **-99 'Missing value'** – used when a user-entered valid answer was expected but none was present without a specified reason (mostly meaning that the respondent skipped an answer).
- **-98 'I don't know'** – a user-entered missing value. This option was provided for all questions.
- **-97 'Prefer not to say'** – a user-entered missing value. This option was provided for all questions.
- **-96 'Routing'** – a value used for routed-out answers. Used if individual respondents were not asked a follow-up question(s) due to their answer to the previous question(s).
- **-95 'Cleaning'** – a missing value created during data cleaning, used if respondents provided contradictory or invalid answers (or when an invalid answer was entered during the data entry phase of the pen-and-paper surveys). The code was used if the value could not be corrected due to the central data-cleaning procedure (based on consultations with the national teams or examination of the value pattern).
- -**94 'Not asked'** – a value used when the whole question was omitted from the survey in the respective country and was therefore not asked during the interviews.
- **-93 'Too young to reply'** – used if younger children were not asked questions deemed to be too complex or sensitive. These were marked with the _oy suffix. However, in some countries, omission for younger children was also applied for other type of questions (see the 'Data Dictionary').
- **-92 'Omitted by error'** – a value indicating that a question was omitted due to a technical or procedural error. The most frequent cause was that older children were given a shortened version of the questionnaire (commonly because the younger and older children were mixed in a class in which data collection occurred).
- **-91 'Different version of questionnaire was used'** – a value used if a respective question was asked with national modification. Mostly used when a country added another answer option to the question or changed the answer scale. This value indicated that there was a national version of the question in the dataset (with a country prefix).

Sampling procedures and fieldwork

The sampling and data collection procedure in the project comprised several methods that were selected based on the main recommended strategies and adjusted for national context.

Sampling strategy

The EU Kids Online Management Group provided guidelines for sampling that aimed to maximize the comparability and representability of each national dataset. The target survey population was defined as children aged 9–17 who use the internet. The recommended minimum sample size (before data cleaning) was 1,000 respondents per country (with a few exceptions, typically for the smaller countries). There were two sampling methods using different sampling points: via households and via schools. Each participating country selected the method depending on available resources and country and cultural context. For sampling, the following criteria were proposed to provide the best combination of representativeness and viability: age of the child, gender of the child, region (usually NUTS 2) and, if applicable, urban/rural areas. The application of all these recommendations was tailored to the national context to provide data that would be representative of the targeted population.

Random-probability sample of households (household sampling)

Countries that used household sampling were Croatia, Estonia, France, Germany, Italy, Lithuania, Norway, Russia and Slovakia. Several approaches in household sampling were implemented. These included a random walk procedure, quota sampling and random recruitment or selection of households from a specific database of addresses.

Random walk procedure: based on the distribution of the surveyed population, specific numbers of households were marked as seed addresses (seed_ID in the dataset), i.e., the starting point for a random walk. The interviewer followed strict predefined path instructions, and tried to contact households selected on these instructions and screen them for eligibility for the research. If contact was not established, the interviewer attempted to contact the residents later. The eligibility criteria were: the willingness to participate of the respondent of a certain age and sex/gender given by the sampling quota. Note: In Russia, the sampling procedure was not able to fully ensure the same probability for residents in different parts of country to take part in the survey, as it only took place in the larger cities. Due to limited accessible information on the distribution of Russian children in the population, Russian data was therefore not weighted (see Annex 1).

Telephone recruitment from a national or other certified register: a list of addresses provided by the register agency was used to contact households from predefined sampling points. Estonia and Norway used sampling via the population register.

Recruitment of households belonging to online panel: in France, recruitment from a national online panel was used. The sample collection was designed to achieve a sample reflecting base population distributions and to ensure equal chances of participation. Criteria for recruiting respondents were regions according to regional distributions (NUTS 2) and the size of the municipalities.

Sampling via school classes (school sampling)

For sampling via schools, the guidelines defined for ESPAD 2015 (i.e., the European School Survey Project on Alcohol and other Drugs) were recommended. The aim was to meet the same methodological and ethical standards defined by this project. For EU Kids Online IV, the general target population was defined as students aged 9 to 17 who were present in the classroom on the day of the survey. Students enrolled in regular, vocational, general and academic studies were included. Those who were enrolled in either special schools or special classes for students

with learning disabilities or severe physical disabilities were excluded. Individual sampling units consisted of school classes. The participation of at least 1,000 respondents per country (with exceptions for a few of the smaller countries) was recommended. Moreover, recruitment of respondents from a wider list of schools and classes was encouraged. These recommendations were formulated in order to adjust for clustering effects within classes and schools.

The countries that used sampling via schools were Flanders, Czech Republic, Finland, Malta, Poland, Portugal, Romania, Serbia, Spain and Switzerland. Note: Flanders and Finland used specific sampling that also precluded the weighting options. Data from Belgium were designed to reflect only pupils from the Flanders region (thus the Belgian contribution for this survey is referred to as Flanders) while also excluding Brussels. Moreover, urban and regional profiles of the surveyed schools differed from population distributions. In Finland, the final sample deviated from population distributions for both the age and region (see Annex 1).

Data collection method

The EU Kids Online IV survey used three methods of data collection, CASI/CAWI, CAPI and PAPI. In all cases, several rules were followed in order to minimize bias due to interview conditions and to comply with the survey's ethical standards (see section 'Ethics'). This especially included consideration of bias caused by the participant not feeling anonymous, which should be diminished by the requirement to ensure the participant's anonymity as much as possible and protection from the influence of outside sources (in households these could generally mean the presence and influence of parents/family; in schools, of teachers or other students).

CASI/CAWI (Computer-Assisted Self-Interviewing/Computer-Assisted Web Interviewing). This method was used in both household and school interviews. Interviewed children filled in the questionnaire on their own on the tablets/notebooks/computers while being instructed by the interviewers. In schools local equipment was used; in household interviewing, equipment brought by the interviewers was used. Even though the children were answering each question on their own, a trained interviewer was present who administered the procedure and ensured there were no ethical, cognitive or technical problems. An exception was France, which used online data collection in households where children filled in their responses alone on their household computers.

CAPI (Computer-Assisted Personal Interviewing). This method was only used in those countries that used data collection in households. Interviewers asked the children each question in its exact wording and marked the answer on an electronic tool. Children were handed the data-collecting tool and filled in the answers on their own for questions that were deemed be sensitive.

PAPI (Paper-Assisted Personal Interviewing). Paper versions of the questionnaire were used, filled in during interviews. The procedure was conducted in the presence of trained administrators. This method was used mostly in countries that used school sampling for their survey. In Estonia and Portugal, this method was used if CASI/CAWI could not be used.

It should be noted that the two general approaches, that is a personal interview with or without an administrator directly marking the answers, differ slightly. The drawback of a direct personal interview includes the higher potential of non-anonymity for the respondent (although a trained interviewer would strive to assure the respondent about their anonymity as much as possible). The benefits include more opportunities to help with the smooth procedure of the interview. During school data collection, while interviewed children may feel more anonymous, the control of the interview may have been decreased.

In countries that applied household data collection, national teams and surveying agencies decided if the use of incentives would be applied. Most countries that used school data collection did not use any form of direct incentives. The exception to this was Switzerland, where teachers from every participating class were given CHF100 (around €90) to use for class-based activities. Most countries that used the household sampling method also used some form of incentive (except for Germany, Lithuania and Russia). Countries that used incentives during their household surveys were:

- Croatia: sweets for interviewed children were bought and provided by the research agency
- Estonia: sweets for interviewed children were bought and provided by the research agency

- France: a small financial donation (€1.8) could be sent to the charity of the participant's choosing
- Italy: small gifts were provided by the research company
- Norway: households were gifted financial incentives of NOK200 (around €20), later changed to NOK400 (around €40)
- Slovakia: small gifts were provided by the research company

An overview of the methodological approach used in each country is shown in Table 2.

Table 2: Overview of the fieldwork

Country	Place of interview	Fieldwork	Method of interview	Survey carried out by	Sampling criteria
CH	School	10/2018 to 01/2019	PAPI	GFS Zürich agency	Age groups (classes), NUTS 3
CZ	School	10/2017 to 02/2018	CASI/CAWI	CZ EU Kids Online team	Age groups (classes), NUTS 2, school size, school type
DE	Household	06/2019 to 07/2019	CASI/CAWI	Ipsos agency	Age categories (9-11, 12-14, 15-17), gender, BIK urban variable, parental education, NUTS 1, household net income
EE	Household	05/2018 to 07/2018	CASI/CAWI*	Turu-uuringute AS agency	Gender, age categories (9-10. 11-12, 13-15, 16-17), NUTS 3, urban type (urban/rural), city districts (in the city of Tallinn)
ES	School	10/2018 to 12/2018	PAPI	CPS Estudios de Mercado y Opinión agency	Age groups (classes), school type (state/mixed), urban type (capitals/smaller towns), edited NUTS 1 classification
FI	School	01/2019 to 04/2019	CASI/CAWI	FI EU Kids Online team	NUTS 2 (or NUTS 3 classification where possible)
FR	Online survey	05/2018 to 06/2018	CASI/CAWI	OpinionWay agency	Age, gender, NUTS 2
HR	Household	09/2017 to 10/2017	CAPI	Ipsos Puls agency	Gender, age, urban, detailed NUTS 2 classification
IT	Household	11/2017 to 12/2017	CAPI	Ipsos agency	Age, NUTS 2
LT	Household	01/2018 to 05/2018	CAPI	Spinter research agency	Age, gender, urban type (urban/rural), NUTS 2
MT	School	03/2018 to 05/2018	PAPI	MT EU Kids Online team and Personal, Social and Career Development (PSCD) educators	Age groups (classes), school type, detailed NUTS 3 classification
NO	Household	06/2018 to 10/2018	CASI/CAWI	Ipsos agency	NUTS 2, classification system KOSTRA (municipality size and economy), gender, age
PL	School	05/2018 to 06/2018	CASI/CAWI	Edbad agency	Age groups (classes), NUTS 2, school type
PT	School	03/2018 to 07/2018	CASI/CAWI*	Intercampus SA agency	Age groups (classes), school type, NUTS 2
RO	School	04/2018 to 04/2019	CASI/CAWI	Romanian Institute for Evaluation and Strategy (IRES)	Age groups (classes), school type, NUTS 2
RS	School	11/2018 to 01/2019	PAPI	RS EU Kids Online team	Age group (classes), regions (as proposed for NUTS 2)
RU	Household	09/2018 to 10/2018	CAPI	RU EU Kids Online team	Age, gender, Federal districts
SK	Household	04/2018 to 06/2018	CAPI	Kantar Slovakia agency	Age categories (9-11, 12-14, 15-17), gender, urban, NUTS 3
VL	School	03/2018 to 11/2018	CASI/CAWI	Institute for Media Studies at KU Leuven	Age groups (classes), school type

* This was a dominant method of interview. In Estonia, out of 1,010 interviews 773 were collected through CASI/CAWI, 115 through CAPI and 122 through PAPI. In Portugal, out of 1,861 interviews, 1,839 were collected through CASI/CAWI and 22 through PAPI.

Ethics

Researching children and young people, and their relationship to online risk, constitutes an ethical challenge that needs reflection and diligence. Children are considered vulnerable informants *per se*. This means that extra care needs to be taken by researchers to ensure informed consent and to avoid potential harm. In this section we address two general considerations. First concerns the questionnaire development, the second the data collection (see Table 3 for details).

Ethical approach in developing the questionnaire

As described in our comparative report[1], researching risk also means acknowledging that what is defined as a potential risk for some can be seen as an opportunity for others. One such example is experiences with sexual messages, where for some young people, under some circumstances, receiving a sexual message from a peer, a girlfriend or a boyfriend can be seen as positive and exciting, while for others such messages may cause distress and potential harm. Consequently, when asking children and young people about their experiences online, we tried, both for methodological and ethical reasons, to avoid normative connotations and guidance. As also described in our comparative report, this led us to extend the questionnaire options to include a wider range of experienced feelings than before, rather than just levels of distress. Children were asked if certain experiences had bothered them or not, without assuming that all children and young people would perceive an experience as problematic and harmful. For this reason, some risk sections also included follow-up questions about any positive reactions and feelings according to what most may perceive as risk-related and/or abusive behaviour.

In addition, individual questions in the EU Kids Online questionnaire not only included the option 'I don't know', but also the option 'I prefer not to say. This is especially important as the questionnaire included some sensitive questions, such as experiences with sexual risks and opportunities, transgressional behaviour and feelings towards family and friends. In order to be comprehensive, the questionnaires were also subject to cognitive testing, both in 2010 and in 2018.[2,3]

Ethics during data collection

In addition to adhering to legislation on the collection of personal data from informants (such as the GDPR and various national legislative provisions), researchers have a duty to safeguard against harm and unreasonable strain, respecting all individuals' human dignity, interests and integrity[4]. This requirement applies to all phases of a research project, from developing a research topic and the appropriate method, via the collection of data and to the reporting of the findings. An important tool for ensuring this is informed consent.

Informed consent means that anyone participating in the research has the right to sufficient information on what they will be asked to do, why, and how the information is to be used, so they can make up their own mind whether or not to participate (self-determination). For younger children and young people, both parental consent and consent from the child him- or herself is needed. As the age of consent varies across Europe, in our survey all

[1] Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*. EU Kids Online. doi: 10.21953/lse.47fdeqj01ofo

[2] Livingstone, S., Haddon, L., Görzig, A., & Ólafsson, K. (2011). *Risk and safety on the internet: The perspective of European children. Full findings from the EU Kids Online survey of 9-16 year olds and their parents*. EU Kids Online, LSE. http://eprints.lse.ac.uk/33731/

[3] Ní Bhroin, N. & Rehder, M.M. (2018). *Digital Natives of Naïve Experts? Exploring how Norwegian children (aged 9-15) understand the Internet*. http://www.lse.ac.uk/media-and-communications/assets/documents/research/eu-kids-online/reports/norway-report.pdf

[4] NESH (The National Committee for Research Ethics in the Social Sciences and the Humanities) (2016). *Guidelines for research ethics in the social sciences, humanities, law and theology*. Oslo.

countries ensured consent from both the parent and the child (unless the child was over the age of consent and parental data was not collected). Data from all informants were anonymised.

Table 3: Details concerning ethical aspects

Country	Consent from parent	Consent from child	Approval of ethical body (if needed)	Incentives
CH	Yes (unwritten)	Yes (unwritten)	Not required	Yes (small)
CZ	Yes	Yes (unwritten)	Yes, MU	No
DE	Yes (unwritten)	Yes (unwritten)	Not required	No
EE	Yes (unwritten)	Yes (unwritten)	Not required	Yes (small)
ES	Yes	Yes (unwritten)	Ethics Committee of Universidad del País Vasco	No
FI	Yes	Yes	Not required	No
FR	Yes	Yes, unwritten	Yes, CNIL	Yes (small)
HR	Yes	Yes (unwritten)	Not required	Yes (small)
IT	Yes	Yes	Not required, but Ipsos has its ethics standard	Yes
LT	Yes	Yes	Vilnius University Faculty of Philosophy Research Ethical Committee	No
MT	Yes	Yes (unwritten)	Yes, Directorate of Education	No
NO	Yes	Yes	Not required, but personal data collection approval via Ipsos (data protection authority)	Yes
PL	Yes	Yes	Yes, AMU ET	No
PT	Yes	Yes (unwritten)	National Committee of Data Protection/Directorate-General for Education	No
RO	Yes	Yes (unwritten)	The Institute of Sociology, The Romanian Institute for Evaluation and Strategy	No
RS	Yes	Yes	Ethics Committee, Department of Psychology, Faculty of Philosophy, University of Belgrade	No
RU	Yes	Yes (unwritten)	Lomonosov Moscow State University	No
SK	Yes	Yes (unwritten)	Not required	Yes (small)
VL	Yes	Yes	KU Leuven Ethical Review Board (SMEC)	No

There are major methodological and ethical challenges associated with mapping risk experienced by children and young people. The countries included in this report collected data by various methods, some at home, some in schools and some using an online panel. While there are different challenges associated with these methods, all countries and teams collecting data paid due attention to the ethical requirements and dilemmas associated with the research. This includes, but is not limited to:

- Whenever possible having a responsible researcher or informed adult/teacher present during the collection of the data who could answer questions from respondents and provide clarification when needed.
- Ensuring informed consent from the respondents. This included emphasizing that participation in the survey was voluntary and that the child could withdraw from the study at any time without any consequences. Typically, the child gave consent to participating in the survey verbally or at the beginning of the questionnaire.
- The child was able to answer all questions anonymously.
- Answers from the child were confidential, to interviewers and parents and/or teachers.
- When collecting data at home, the child was in a separate room with no influence from the parents.
- When collecting data, all participants should have been given information in their own child-friendly language where they could find more information regarding the topics covered in the survey.

Data entry and data management

Data management was conducted by the EU Kids Online Data Management Group, which consisted of the following researchers: Rostislav Zlamal (CZ), Hana Machackova (CZ), David Smahel (CZ), Kjartan Ólafsson (IS), Katarzyna Abramczuk (PL) and Elisabeth Staksrud (NO). Basic procedures concerning data management were as follows:

- The EU Kids Online Data Management Group provided a predefined unified SPSS matrix for all teams using coding described in section 'EU Kids Online IV questionnaire'.
- The Data Management Group also prepared a template for the technical report that was provided to the national teams.
- The national teams inserted their national data into this predefined SPSS matrix. They prepared their national technical reports that comprised detailed information about the methodology used, as well as the data entered into the matrix, including lists of variables used. The national teams sent their data and technical report to the Data Management Group for controls and data cleaning.
- Two members of the Data Management Group, namely Rostislav Zlamal and Katarzyna Abramczuk, worked on data quality control and data cleaning (see below). The Data Management Group members cooperated with the national teams to reach maximum data quality. After all data quality controls and data cleaning, the final version of data was sent to the national team.
- This final version was merged into the international database.

Data entry and data cleaning at national level

All the national teams from EU Kids Online were asked to provide their data in the predefined SPSS matrix. The nature of the quality of the data was related to the national methods of data collection and sampling. The first level of data cleaning was typically done at national level. All datasets collected by the professional agencies were first cleaned by the relevant professional agency. Similarly, most national teams also performed the first data cleaning in the case of data collection through schools. Cleaning procedures were different across the national teams, so the national researchers were asked to describe their data cleaning procedures in as much detail as possible in the national technical report.

Data cleaning and quality diagnosis

The Data Management Group applied quality checks on the national data focused on controlling the logical structure of the questionnaire, missing data, possible missing variables or other problems with the datasets. The Data Management Group developed a set of scripts (in R and SPSS) for the quality diagnoses that were systematically applied to each dataset. These were divided into the following basic categories:

Quality checks diagnosing the fit with the SPSS matrix

The first step was to check whether the national data fitted or deviated from the predefined data matrix. The scripts in SPSS were used to check the list of variables used, the ranges of defined values and deviations from the predefined labels of values. Any possible errors in the structure of the individual datasets (e.g., omitted core questions, different defined values or a large amount of system-missing values) were discussed with the national teams and corrected after discussion. However, the national datasets also included deviations that could not be corrected. These are listed in Annex 1.

Manual checks for deeper exploration of data and concurrence with the technical report

These procedures aimed to check the accuracy of demographic sampling and auxiliary variables on the information provided by the national technical reports. These included checking codes for geographical clusters (NUTS), the exact calculations of age, differences between older and younger groups of children, possible conflicts between interview dates, checks and crosschecks of class and school sizes, interview lengths and the provided country variables. Additional manual checks verified that there should not be duplicate observations or identical IDs. The national technical report was also checked with regard to adherence to general EU Kids Online IV principles.

Handling various types of missing values

The guidelines and predefined matrix included a specific system for the coding of different types of missing values (see section 'Missing values'). The Data Management Group provided procedures to check the consistency of the missing values coding and whether the applied system was in accordance with the EU Kids Online IV guidelines. The most common erroneous coding concerned the categories of -99 ('missing value'), -96 ('routing'), and -94 ('not asked'), which were often interchanged. A set of scripts was developed to check the accuracy of the missing values that were used. Moreover, two new missing values were added in the data: -93 ('too young to reply') and -92 ('omitted by error'). These two values were created for variables that should have been (in some countries) asked only of older children. Value -93 was used to indicate missing data in younger children who were not asked a respective item (typically modules and variables with _oy suffixes). Missing value -92 was used when older children were not asked questions for older respondents by accident. This error mostly occurred in school-based surveys in which the same type of questionnaire (i.e., either for younger or older children) was administered to a whole class, based on the lower age range in class.

Diagnostics of missing values

The final step in data checking included procedures that aimed to check overall data quality with respect to the ratio of missing values. Due to technical or other errors, several types of missing values and blank observations were present in each national dataset. The Data Management Group set up a threshold for ratios of the system-missing values. Specific scripts were developed to check missing values in some of the key core questions together with the overall level of missing values for each observation. The observations with a high level of missing values were removed or later weighted by zero (see section 'Weights creation').

During the data quality diagnostics, any possible deviations from the target population were searched for. Observations outside of the surveyed range were handled in a similar way as low-quality observations (weighted by zero). These procedures included cleaning of observations if the children were outside the age range 9–17 and when children were non-internet users.

In summary, the main reasons why observations were weighted by zero were: overall low quality of the whole observation (more than 50% of system-missing values), large amount of missing values in key core variables, children outside of a desired age range, and children not using the internet.

Data merging

As a final step, all national datasets were merged into one data file, based on the original version of the database. In data merging, specific cases of national data differences had to be solved. The most common were when the national question differed slightly from the original question. In these cases, special national variables were created and inserted into the international database. Individual country-specific variables and reasons why they differed from the EU Kids Online template are given in Annex 1. Furthermore, many national teams used questions that were specific for their national surveys. These were not entered into the international database.

Data weighting and design effects

Use of weights in the dataset

A sample may cover segments of the target population in proportions that do not match the proportions of those segments in the population itself. For example, the share of girls in the sample can oftentimes be larger than the share of boys, while in the population it is the number of boys that is larger than the number of girls. The sample totals may differ from the population totals for a number of reasons. They may result from the sample design, non-response or random fluctuations. Returning to the example, boys may be less common in the sample if they are less willing to participate in surveys than girls. This is a non-response issue. Taking another example, the sample may be designed in a way that ensures that there is an equal share of children from every region to enable reliable regional estimates. If the regions are of various sizes this will lead to a mismatch for the geographical region variable at a national level. To correct for this type of mismatch every national dataset includes a weight variable.

The weights available in the dataset are *country-specific*. They are meant to correct for bias caused by unequal probabilities of selection and varying response rates across different types of respondent within each country. These can be used for country-by-country analysis, and for analysis looking at any single country. For example, we can compare two or more countries in terms of frequency of some phenomena. It is not possible, however, to generate European-level distributions, as there is no European-level weight in the data. Countries with various population sizes participated in the study. The samples in individual countries were chosen to ensure valid representation of the country population. This means that countries with very different population sizes, such as Germany and Lithuania, had a similar representation in the data – around 1,000 observations. Due to this mismatch it was not possible to generate European-level distributions.

For three countries – Flanders, Finland and Russia – it was impossible to create weights of high quality (see section 'Sampling procedures and fieldwork'). In these cases all the observations were assigned weights equal to one, which is the equivalent of not weighting them at all. Caution is advised when drawing population-level conclusions for these three countries.

Consistently with the guidelines developed by the EU Kids Online network in the 2010 study, these weights can be applied to make *descriptive statistics* representative of the population. For statistical significance testing weights should not be applied to avoid biased standard errors.

The effect of weighting can be large whenever a very small subgroup of the sample is considered. In this case even small differences in weights may dramatically change the distributions of the analysed variables. For this reason it is not advisable to use weights when considering rare phenomena. This may be particularly important for routed questions.

Context information

In general the weights in the dataset are *raking weights*. These aim to improve the relationship between the sample and the population so that the marginal totals of the adjusted weights on specified characteristics agree with the corresponding totals for the population. Creating raking weights requires specifying the characteristics whose distributions are to be corrected. A weight value is assigned to each respondent in a way that ensures that the weighted distribution of the sample is in very close agreement with the distribution of the chosen variables in the population. One of two possible types of weights was created: RIM weights or post-stratification weights.

Weights creation was carried out either by a member of the EU Kids Online Data Management Group or the agency tasked with data collection. The Data Management Group conducted a quality diagnostic for the weights prepared by the agencies (details are given below). Furthermore, if there were any changes in the final sample size

due to the removal of low-quality observations, the agency was contacted and asked to adjust their weights. The Data Management Group then checked the adjusted weights.

Additional weights were available for three countries. For Italy and Norway these were weights created by the agency tasked with data collection. In the case of Romania, these were weights that could be used when analysing responses to the module questions. As a result there were three weighting variables in the dataset:

weights – the main weighting variable. This should be the default choice when conducting a descriptive analysis. It was used in the comparative report of EU Kids Online 2020.

weights_modules – this is identical with the 'weights' variable for all countries except for Romania. A technical error during study implementation in Romania led to omission of the core questionnaire for 53 observations, although module questions were still asked. The 'weights' variable treats these 53 observations as missing and assigns a weight of zero to them. The 'weights_RO_modules' variable treats them as regular observations with non-zero weights. This variable should be applied for module analysis.

weights_agency – this is identical with the 'weights' variable for all countries except for Norway and Italy. In these two cases it contained values of the weighting variables provided by the agency conducting data collection. The agency-provided weights were also raking weights but were created using a wider set of variables including information about parents, namely, their education and for Norway, gender. This information was not part of the sampling frame and significantly decreased the quality of the weights. For the sake of consistency with the other datasets and the sampling frames the weights were redone. However, the original agency weights are still available in the dataset.

It is common practice, as was the case in the 2010 EU Kids Online survey, to approach weighting in stages. In the process design weights are computed. Their goal is to adjust directly for the unequal probabilities of selection during sampling. Due to difficulties in identifying the exact sampling frames, the statistical information used in the process of sampling, and non-response rates of the weights created by the Data Management Group in this wave of the study, could not be prepared this way. The only exception was the data for the Czech Republic where the sampling weights could be computed and used as the basis for raking weights.

Weights creation

All the weights in the dataset were correct for the distribution of three variables: gender, age/grade and geographical region. Official national statistics were used as reference distributions. It is important to keep in mind that this data refers to the whole population of children, non-internet users included, while the target population of the study was internet users. There is no available data on the population of children aged 9–17 who use the internet by country. However, the estimates ran in 2010 indicate that the vast majority of European children can be classified as internet users. The estimated share of children online in most countries in 2010 varied between 78% and 98%. These shares are likely to be higher in the age of mobile technologies.

The age variable used for weighting was computed so as to match the available data as closely as possible. The age variable in the dataset corresponds to the exact age at the time of the study while the official statistics refer in most cases to the age at the end of a year. The latter age was used to compute weights. In some cases estimates were made to assess the share of children of a given age that belonged to the sampling frame. In particular, when school sampling was used the share of youngest and oldest children available for sampling was estimated. For Serbia official statistics regarding grade rather than age were used. In some cases the age/grade variable was grouped. For Poland this was done to avoid excessively high or low weights for some observations, which could have resulted from low availability of one age group. This low availability was a consequence of a reform in the educational system that was in progress during data collection.

The level of regional classification that was used for weights creation differed depending on the country. It varied with the country size and the sampling scheme used. It could be NUTS 1, NUTS 2 or NUTS 3 classification valid at the time of the study. Whenever it was needed and justified (e.g., due to the sample selection method) additional variables were also included in the raking process. These were often variables used to define strata in the sampling process such as the 'urban' variable or the 'type of school' variable.

Table 4 presents information on the source of weights and distributions taken into account during weights creation. It refers only to the main weighting variables. Differences between this weight and the remaining two weighting variables are described in the preceding section.

Table 4: The source and variables used for creating the main weighting variable in the dataset

Country	Source of weights	Variables used for weights creation			
		Age	Gender	Region	Other
CH	EUKO DMG	Age	Gender	NUTS 2	Type of agglomeration (rural, urban, suburban)
CZ	EUKO DMG	Age	Gender	NUTS 2	Urban variable, school type, school size
DE	Agency	Age categories (9–11, 12–14, 15–17)	Gender	NUTS 1	BIK variable (i.e., classification of urbanization), education of parents, net household income
EE	Agency	Age	Gender	NUTS 2	Urban variable, language
ES	EUKO DMG	Age	Gender	NUTS 1	-
FI	N/A	-	-	-	-
FR	EUKO DMG	Age	Gender	NUTS 1	-
HR	EUKO DMG	Age	Gender	NUTS 3	-
IT	EUKO DMG	Age	Gender	NUTS 2	-
LT	EUKO DMG	Age	Gender	NUTS 3	-
MT	EUKO DMG	Age	Gender	NUTS 3	-
NO	EUKO DMG	Age	Gender	NUTS 2	-
PL	EUKO DMG	Age (15–16 as one category)	Gender	NUTS 1	School type
PT	EUKO DMG	Age	Gender	NUTS 2	-
RO	EUKO DMG	Age	Gender	NUTS 2	School type
RS	EUKO DMG	Grade	Gender	NUTS 2	Urban variable, school type
RU	N/A	-	-	-	-
SK	Agency	Age	Gender	NUTS 3	Urban variable
VL	N/A	-	-	-	-

* EUKO DMG = EU Kids Online Data Management Group.

After the initial raking, weights can vary greatly. Some respondents may have extremely low or high weights relative to most of the other respondents. If these weights are used, it would mean that responses of some children would weight, for example, 20 times more than responses of other children. Excessive variation in weights values leads to inflated sampling variances of the survey estimates. To solve this problem weights were trimmed to ensure reasonable weight values.

For 464 observations the main 'weights' variable takes on a value of zero. It was assigned to those observations that did not fulfill all the requirements of Kids Online IV study but can still be interesting to look at in some specific analyses. The zero-weighted cases include:

- Children outside of the target age group
- Observations from Romania for which the core questionnaire was omitted (see above)
- Low-quality observations (see section 'Data cleaning and quality diagnosis')
- Children who were deemed non-internet users based on their answers to QB5 items

Intraclass correlation and design effects for school samples

One of the biggest methodological challenges related to the EU Kids Online IV study stems from the fact that sampling methodology varied across countries. While in some countries random samples of households were used, others relied on school sampling. Data gathered using these two methods can be difficult to compare because of the clustering issue. Typically in school sampling whole classes are selected at once. As children attending the same class, or even the same school, are likely to be more similar to each other than a random group of children, it may lower the overall quality of the sample in terms of variance and consequently, the effective sample size. To estimate the size of this problem design effects for school-based samples were computed.

Design effects are 'the ratio of the sampling variance for a statistic computed using a [particular design] divided by the sampling variance that would have been obtained from a [Simple Random Sample] of exactly the same size'.[5] In the current case the design effect indicates the loss of precision in survey results derived using school sampling compared with the reliability of results derived using a Simple Random Sampling method. For data collected using cluster sampling design effect can be computed on the basis of intraclass correlation (ICC). It equals $\rho(m - 1) + 1$, where ρ is the intraclass correlation coefficient and m is the average number of consultations per cluster.[6] The intraclass correlation coefficient is a measure of the homogeneity of elements within clusters and has a maximum value of +1 when there is complete homogeneity within clusters, and a minimum value of $-1/(m - 1)$ when there is extreme heterogeneity within clusters.

To estimate intraclass correlations, IBM SPSS software was used. Schools were treated as clusters. Since each statistic in a survey has its own design effect, a set of variables was selected to perform computations. These variables were selected purposively to cover a range of different types of question, and therefore to give an indication of the range of design effects that may apply to different types of questions. These were also selected to cover some of the key measures of interest from the survey (including internet use, parental monitoring and knowledge, exposure to risks online and child self-sufficiency) and to provide an indication of the psychological profile of children from different sampling points. The variables used to compute the design effects are shown in Table 5.

[5] Groves, R.M. (2004) *Survey methodology*. Hoboken, NJ: Wiley.
[6] Kish, L. (1965) *Survey sampling*. New York: John Wiley & Sons, Inc., p. 162.

Table 5: Variables used to calculate design effects related to clustering in school samples

Variable name	Variable description
Highly clustered variables	
c_QB5a	How often do you go online or use the internet using the following devices? A mobile phone or smartphone
c_QB7	About how long do you spend on the internet during a regular weekday (school day)?
c_QC3b	How often have you done these things ONLINE in the past month? I used the internet for schoolwork
c_QC3h	How often have you done these things ONLINE in the past month? I visited a social networking site
c_QF11	Have you EVER had contact on the internet with someone you have not met face-to-face before?
c_QF30	In the PAST YEAR, have you EVER SEEN any sexual images?
Other variables	
c_QA9	Here is a picture of a ladder. Think of this ladder as representing where people stand in your country. PLEASE TICK THE BOX WHERE YOU THINK YOU AND YOUR FAMILY ARE
c_QF01	In the PAST YEAR, has anything EVER happened online that bothered or upset you in some way (e.g., made you feel upset, uncomfortable, scared or that you shouldn't have seen it)?
c_QF20	In the PAST YEAR, has anyone EVER treated you in such a hurtful or nasty way?
c_QF28	In the PAST YEAR, have you EVER TREATED someone else in a hurtful or nasty way?
c_QF50c_oy	In the PAST YEAR, have you seen online content or online discussions where people talk about or show any of these things? Ways to be very thin (such as being anorexic or bulimic, or 'thinspiration')
c_QF50e_oy	In the PAST YEAR, have you seen online content or online discussions where people talk about or show any of these things? Their experiences of taking drugs
c_QF60a	In the PAST YEAR, has any of the following happened to you on the internet? Somebody used my personal information in a way I didn't like
c_QA10a	How true are these things of you? I get very angry and often lose my temper
c_QA18a	How true are these things of you? I do dangerous things for fun
c_QI2a	How true are the following things about your family and home? When I speak someone listens to what I say
c_QI4a	When you use the internet, how often does your parent or carer do any of these things? Encourages me to explore and learn things on the internet
c_QI13	Do you ever ignore what your parent or carer tells you about how and when you can use the internet?
c_QJ1a	Please say how much you agree or disagree. I feel like I belong in my school
c_QK1a	How true are the following things for you? My friends really try to help me

Some of the selected variables are more and some are less prone to clustering within schools. After computing intraclass correlations for each of these variables, a group of highly clustered variables was identified. They correspond to questions to which pupils within the same school tend to give similar answers. These include such phenomena as school-related online activities and social media usage. The design effect for such variables is generally higher.

Table 6 shows the results by country. It gives the unweighted sample size for each country, i.e., the actual number of interviews conducted, as well as the approximate intraclass correlation, the approximate design effect and the approximate effective sample size for both the highly correlated variables and other variables. It gives a general idea of the importance of the clustering effect in school-based samples.

Table 6: Design effect related to school sampling

	Actual sample size	Highly clustered variables			Other variables		
		Approximate intraclass correlation	Approximate design effect	Approximate effective sample size	Approximate intraclass correlation	Approximate design effect	Approximate effective sample size
VL	1,392	0.05	5.74	368	0.04	3.73	504
CH	940	0.35	5.72	176	0.08	1.98	531
CZ	2,824	0.23	7.63	378	0.06	2.68	1,169
ES	2,872	0.17	7.88	395	0.05	2.79	1,272
MT	1,232	0.18	10.99	125	0.03	2.65	660
PL	1,168	0.21	7.39	162	0.07	2.92	427
PT	1,861	0.17	3.70	553	0.04	1.62	1,399
RO	868	0.16	13.60	70	0.00	1.16	807
RS	1,150	0.30	6.21	200	0.07	2.18	595

The easiest way to interpret the design effect is with reference to the effective sample size calculated as: actual sample/design effect. The effective sample size shows the amount of confidence we have in the reliability of our figures, after adjusting for the impact of the survey design. It varies across variables. Table 6 clearly shows that for highly clustered variables caution is recommended while in other variables sample efficiency is more reasonable.

Effective sample size

To ensure the sampling and weighting quality, we computed effective sample size and effective sample size proficiency for each country, excluding Flanders, Finland and Russia (for the reasons mentioned above). Another reason to compute the effective sample size proficiency was to check if the clustering effect of school sampling did not lower the overall sample size.

For calculation, we used Kish's effective sample size. This method uses weights and their individual values to compute the impact of weighting. In other words, it measures how far from ideal a distribution weighted sample really is and how small the weighted sample could be, if sampling was carried out in a perfect way but still provided the same level of measurement quality. The base formula of Kish's effective sample size is as follows:

$$n_{eff} = [\Sigma\, \omega_i]^2 / [\Sigma\, (\omega_i^2)]$$

The effective sample size proficiency was then calculated from an effective sample size and total number of weighted observations. This provides an overview of how small the effective sample size is in reality in contrast to the original national dataset. It is noteworthy that only the sample of weighted observations was used in this computation. This means that any observation weighted with a value of zero (e.g., too young or too old children, low-quality observations, non-internet users; see (see section 'Weights creation') are not part of the effective sample size proficiency calculation. Generally, the number of weighted observations and weighted sample N should be quite similar.

There is no general rule on how low the minimal effective sample size should be. The EU Kids Online Management Group decided that the desired total N of each sample would be 1,000 observations to ensure the minimum of 500 units in the effective sample size. This was deemed to be eligible to take part in the basic exploration analysis of comparative report[7].

[7] Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*. EU Kids Online. doi: 10.21953/lse.47fdeqj01ofo

Table 7: Effective sample sizes

Country	Effective sample size	Effective sample size proficiency
CZ	2,328.68	82.43%
DE	1,007.10	96.47%
EE	930.14	92.18%
ES	2,448.82	85.27%
FR	889.51	92.46%
HR	938.35	92.27%
CH	718.66	76.45%
IT	795.84	79.35%
LT	986.35	97.95%
MT	960.38	77.95%
NO	928.44	97.12%
PL	972.99	83.30%
PT	1,403.86	75.44%
RO	532.04	61.29%
RS	1,010.99	87.91%
SK	946.14	97.64%

Computed values served as a check for any potential errors during sampling and weighting procedures. Initial values of effective sample size proficiencies of datasets from Norway and Italy flagged up problematic calculation of original weights. These had to be made anew. The problematic aspect of original weights was found in implementing parental distributions into weighting procedures. Original weights of datasets from Norway and Italy were imputed into an additional non-default set of weights.

Most countries were able to collect enough observations to surpass a threshold of 1,000 weighted observations. However, due to the strict weighting conditions and the labelling of unfit observations with weights of a value of zero (too young or too old children, low-quality observations, non-internet users), five countries ended up with total weighted N lower than was recommended: France (962), Switzerland (940), Norway (956), Romania (868) and Slovakia (969). Three of these countries (France, Norway and Slovakia) managed to collect weighted datasets with relatively high effective sample size proficiency (more than 90%). The other two (Switzerland and Romania) collected weighted datasets with lower effective sample size proficiency, but the overall effective sample size still surpassed the 500 threshold.

However, it is still worth noting that Kish's effective sample size offers a calculation for the whole dataset. In reality the effective sample size could differ a little in relation to individual variables and more so, if there was analysis to moderate the datasets into smaller and more specific categories of observations. The overall computation of Kish's effective sample size was able to check the quality of sampling and weighting procedures, but we recommend computing an individual variable-dependent effective sample size for more sophisticated analysis that could divide observations into smaller and more specific categories.

Annex

Annex 1: Country specifics

General considerations

- **Multiple-choice answer sets** offered conflicting combinations of possible answers. Answers 'Prefer not to say' and 'I don't know' were often in conflict with valid values of sets. In such cases the valid values were recoded in -95 ('cleaning').
- Values of **school_type** variable were not unified. The nature of the individual categories is country-dependent.
- Due to low interest in the participating countries it was decided to omit module **eHealth** (M4) from the comparative international dataset.

Country specifics

Flanders

- The survey in Belgium was only carried out in **Flanders**.
- The country was labelled as 'specific' and was not weighted. Urban and regional profiles of surveyed schools differed from population distributions. Due to some concerns about weighting and its impact on overall data quality, the weighting procedures were not carried out.
- Core item **c_QB5h** was not used.
- Those aged 12 and younger children were underrepresented or not collected at all. We recommend analysing the sample for the age range <13;17>. This corresponds with the secondary level of the education system.
- In the original dataset 21 observations had the wrong coding for item **c_QF70a_oy**. This had to be recoded to -99 ('missing values').
- All questions between **m1_14a_rt** to **m1_14i_rt** were merged into one question with only one possible answer. This new item was coded as **VL_m1_14a_thru_i_rt**.

Croatia

- Croatia had a greater list of unused core items: **c_QA2a**, **c_QA9**, **c_QE1h_oy**, **c_QF13_rt2**, **c_QF24_rt2**, **c_QF32_rt**, **c_QA12d**, all of **c_QA16a-j_oy**, **c_QA21c**, **c_QA21d**, **c_QA21e**, **c_QA21f**, **c_QA21i**, **c_QI2a**, **c_QI2b**, **c_QI2c**, **c_QI3a**, **c_QI3b**, **c_QK1a**, **c_QK1b** and **c_QK1c**.
- Translation of the dataset was not done in a unified way, and translation itself was not tested.
- Croatia was one of the countries that did not use the item **c_QA2a** (month of birth). Instead, the national team asked directly about **age**. Item **c_QA2b** (year of birth) was then recoded by the national team as an expected year cohort based on the exact age.

Czech Republic

- Items **c_QA18a**, **c_QA18b**, **c_QA21c**, **c_QA21d**, **c_QA21e**, **c_QA21f** and **c_QA21i** were not asked of younger children by mistake. Missing values were coded as -92 ('omitted by error').
- There were no **class_size** data for few classes due to technical error.
- Module item **m1_8h_rt** was not asked.

Estonia

- All core items were used.
- All items of **op_QC4**, **op_QF03** and **op_QD4** were asked in different formats with different sets of values. However, it was possible to recode values in the matrix format.

Finland

- The country was labelled as 'specific' and was not weighted.
- 293 observations had an unknown month of interview – system missing value -99 ('missing value') in an item **int_month**.
- Individual ages of 10 and younger were underrepresented or not collected at all. Children aged 17 were also highly underrepresented. We recommend analysing the sample for the age range <11;16>.
- Core item **c_QA2a** was not asked.

France

- Items **FR_ec_QA6a** to **FR_ec_QA6f** differed from the original template due to the new value 7: 'Almost all the time'.
- France applied a longer list of languages and added items **FR_ec_QA7h**, **FR_ec_QA7k** and **FR_ec_QA7l**.
- All items of **op_QD4a_rt** to **op_QD4e_rt** were asked as one single option question instead of one question with multiple possible answers. These items were thus recoded into country-specific item **FR_op_QD4a_thru_e_rt**.
- Core items **c_QA21f** and **c_QA21i** were not asked.
- Survey was sampled by NUTS 2, but for weights joined NUTS 1 distributions were used.

Germany

- Core items **c_QA2a**, **c_QA2b** and **c_QF11** were not asked.
- The dataset does not include age variables **c_QA2a** (month of birth) and **c_QA2b** (year of birth). Instead, the children were asked their age directly during the national survey.
- Items **c_QF12_rt1** and **c_QF13_rt2** are in the EU Kids Online 2018 template matrix under routing condition from item c_QF11. However, this item was omitted from the national survey. This means that these two items needed to be marked as specific items for Germany (**DE_ c_QF12_rt1** and **DE_ c_QF13_rt2**).
- Standard age categorization of the national dataset regarding younger/older interviewed children is 9–11 and 12–17. However, items **c_QF50b_oy** and **c_QC3a** divide the sample into age ranges 9–14 and 15–17.
- Items **m3_7a**, **m3_7b**, **m3_7e**, **m3_7g**, **m3_7h**, **m3_7k**, **m3_7l** and **m3_7m** include a high amount of 'missing values' (-99). Reasons for why this occurred are currently unknown. A possible explanation may include technical error or use of an unspecified filter.

Italy

- The Italian dataset omitted routing for **m1_5**. In normal circumstances this item should have been routed out of **m1_4≠1**.
- Core items **c_QF23a_rt2**, **c_23f_rt2** and **c_QF24_rt2** were not asked.
- Extended core item **ec_QB5f** was not asked even though the rest of QB5 items were. During cognitive testing the children did not understand the meaning behind 'internet-connected toys'.

- Due to an error there are 47 missing values (-99) out of 57 expected-valid non-missing values in all items of **QF29**.
- **Seed_ID** has only one value: 1. According to the national technical report there should have been 100 PSUs (primary sampling units) + 100 reserves. Given that the seed_ID variable was coded under only one value, it is impossible to determine the exact number of sample points from the national technical report and dataset alone. After consultation with the national team and the agency it was deemed impossible to input seed_ID in its correct form.
- The Italian **urban** variable does not include the lowest value ('Village – below 1,000 inhabitants'). This is due to the fact that the involved agency used different forms of urban classification for Italian sampling and fieldwork. This means that in reality the second lowest value includes observations from both the villages and towns.

Lithuania

- All core items were used.
- Some items of QF10 followed item c_QF11 in the questionnaire and were thus subjected to the extra routing condition. These items are coded as **LT_c_QF10b**, **LT_c_QF10c** and **LT_c_QF10e**.

Malta

- All core items were used.
- The survey was designed for the age range <9;16>.
- Primary independent schools were omitted from the survey, as this category is small and hard to reach.
- No data for class and school sizes are available.
- Some systematic errors occurred during the manual data entering from inputting the responses to the PAPI questionnaires into electronic format. These caused a low level of system-missing values -99 'missing value' to be present throughout the whole dataset.

Norway

- Most other countries defined the younger category of children within the age group of 9- to-10-year-olds. Some 11-year-olds were included in the category of **youngers** and were given a shorter version of the questionnaire. This applies to 66 out of 104 11-year-olds.
- The Norwegian dataset had some changes in wording and lists of offered categories of values. Some of these cases are highlighted as variables with the **NO_** prefix. Edited variables include:
    - **NO_c_QA9 and NO_c_QH1** – both are 'ladder' questions (originally c_QA9 and c_QH1) that were supposed to have 11 cells (values), but Norwegian versions of both of these questions had 10 cells (values).
    - Items of **NO_op_QA17a-j** – these were under an added routing. Individual items of QA17 were routed out if given items of QA16 were answered as 1 ('Never').
    - Item **QF14** – under an additional routing condition. QF14 was routed out if QF13 was answered as 2 ('I was not happy or upset').
    - All used items from **QF22 to QF27** – these were under an additional routing condition. The routing was also applied if QF21b was answered 2 ('A few times').
    - Due to an error there are 20 missing values (-99) in **op_QF06c_rt** instead of valid non-missing values.
    - Due to an error there are 4 missing values (-99) in **ec_QF25_rt3**, all in cases where in c_QF24_rt2 1 was answered ('I was not upset').
    - Core items **c_QK2a**, **c_QK2b** and c_**QK2g** were asked only of older children.

Poland

- The **age** group of 15-year-olds is highly underrepresented. For the purpose of weighting it had to be merged with the age group of 16-year-olds.
- **Ec_QA8a-g** items had no -97 or -98 options available. If pupils did not want to answer 'yes' or 'no', they had a simple option to skip these items. Originally, these had a form of system-missing values, but after discussion of the EU Kids Online Data Management Group it was changed into user-entered missing value of -99.
- Values of all items of **PL_op_QI8a-f** (originally op_QI8a-f) were merged into two categories: 'sometimes/often' and 'never/hardly ever'.
- Values of items **PL_op_QC5a-i** (originally op_QC5a-i) do not include the value 'At least every month'. This value was omitted from the survey by mistake.
- Some of the **school_IDs** may be coded in wrongly. Several cases of schools with low numbers of observations (ranging from 1 to 7) were found.
- Due to the educational reform that took place during data collection, the **school_type** of middle school is underrepresented.
- Survey was sampled by NUTS 2, but for weights joined NUTS 1 distributions were used.

Portugal

- All core items were used.
- Villages were omitted from the survey (there are no cases of value 'Village – below 1,000 inhabitants' of the **urban** variable).

Romania

- Due to a technical error in the interview tool, 65 pupils were handled as non-internet users, and were thus not able to fill in considerable parts of the main questionnaire. It was decided to include these observations in the merged dataset, as their responses for modules and complementary items in the main questionnaire are still valid. These children are included in the **weights_RO_modules**.
- Core items **c_QA3j**, **c_QB5h, c_QD2f** and **c_QA21i** were not asked.

Russia

- This country was labelled as 'specific' and was not weighted.
- The survey only took place in cities (Value 'City – more than 100,000 inhabitants' of the **urban** variable).
- Individual ages of 10 and younger were effectively underrepresented or not collected at all. We recommend analysing the sample for the age range <11;17>.
- Core items **c_QF30**, **c_QF31a_rt**, **c_QF31b_rt**, **c_QF31c_rt**, **c_QF32_rt**, **c_QF40_oy**, **c_QF45_oy**, **c_QF46a_rt_oy**, **c_QF46b_rt_oy**, **c_QF46c_rt_oy** and **c_QF47_oy** were not asked.

Serbia

- All core items were used.
- Villages were omitted from the survey (there are no cases of value 'Village – below 1,000 inhabitants' of the **urban** variable).
- There is no NUTS classification applied yet. Regions currently proposed for NUTS 2 were used for sampling and weighting.

Slovakia

- 21 observations were routed out in **op_QF16_rt1** even though the routing from c_QF12_rt1 should not have been applied. These were recoded in -95.
- Routing for items **op_QF27i_rt2** and **op_QF27j_rt2** did not allow for the entering of valid answers of six children who answered item c_QF21b_rt1 with 2 ('A few times') and one child who answered 3 ('At least every month'). Missing values -96 were recoded in -95.
- Core item **c_QB5h** was not asked.
- The data template includes missing values for **SK_op_QD4a-e_rt** items but none were entered. This is due to an error during which the option of user-missing values was not offered. Moreover, the values of user-entered valid answers follow the older format template of QD4 items used by mistake. These values include <1;4> ('Never', 'Sometimes', 'Often' and 'Always') instead of <0;1> ('No' and 'Yes').
- Due to removing some problematic observations, the final sample of data dropped below the recommended 1,000 minimum. The final sample is **N=969**. However, the effective sample efficiency of the weighted Slovakian sample is rather high (97.63%). This means that the effective sample size of the sample is around 946 observations.

Spain

- Core items **c_QA3i**, **c_QA3j**, **c_QF04i_rt**, **c_QF04j_rt**, **c_QF05l_rt** and **c_QF05m_rt** were not asked.
- The survey was designed for the age range <9;16>.
- Villages were omitted from the survey (there are no cases of value 'Village – below 1,000 inhabitants' of the **urban** variable). Large cities (value 'City – more than 100,000 inhabitants') are overrepresented.
- During data collection some NUTS 2 regions were not fully reached, thus NUTS 1 classification was used for weighting.

Switzerland

- Core items **c_QA3i**, **c_QA3j** and all of **c_QF04a-j_rt** were not asked.
- Items **c_QF29a-c_rt** were subjected to a technical error that caused invalid data for observations that were not supposed to be filtered by the previous filter.
- Survey was sampled by NUTS 3, but for weights joined NUTS 2 distributions were used.

Annex 2: Key data usage rules

Key recommendations for data users. Please ALWAYS check the following:

- **Variables:** What type of variable did you use? Here is a system to use which will guide you in assessing which data you can use. Most importantly:
  - **Prefixes**:
    c_ – core question, should be used by all countries (see 'Country differences' for exceptions)
    ec_, **op_** – extended core or optional, many countries do not have data
    m_ – modules, used just in some countries, many do not have data, require specific weights
  - **Suffixes:**
    _rt – routed out (i.e., answered only by those with the relevant experience, such as only by those victimized about a type of victimization) – there is A LOT of missing data, check which data you need, check missing data under missing value -96
    _oy – optional for younger children, for younger children who do not have data in these variables in at least one or more countries (see 'Age')
  - **Country differences:**
    Some countries omitted and/or altered even core questions. Such altered questions were labelled by a prefix (e.g., **NO_**), the data were not included in the original variable (there is a missing value of -91) and it must be considered if it is possible to merge the variables.

Rule 1: Check the type of variable according to the prefix and suffix and consider missing data. Look at the variable in the 'Data Dictionary' and check if the country has valid data.

- **Age:** The data across countries differ in regard to the age of the respondents. There are two reasons for this:
  - Base sampling: Different countries sampled different ages (e.g., Flanders only sampled those aged 12+, Finland and Russia only those aged 11+, Malta did not sample 17-year-olds, etc.).
  - Younger children category: Countries used different thresholds for **_oy** variables + sometimes did not ask younger children variables not labelled **_oy**.

Rule 2: Always check carefully which age groups are available for analysis within selected countries. Always use age filters (see the next section).

- **Selection of data:** The dataset also includes data from children beyond the age range 9–17, and lower-quality data (many missing values). We highly recommend not using these in most analyses. Use the following filter variables, considering the age range in countries of interest:

  - **Key_filter1:** filters out observation beyond 9–17 and low-quality data (several countries don't have all the data!)
  - **Key_filter2:** filters out observation beyond 9–16 and low-quality data (several countries don't have all the data!)
  - **Key_filter3:** filters out observation beyond 12–16 and low-quality data (all countries included)
    If using another age range, combine with Filter_data_quality, which excludes low-quality data.
  - **Filter_EUKO2020:** selects observations that were used for the EU Kids Online 2020 comparative report – sufficient data quality and age range 9–16. It differs from Key_filter2 by omitting the youngest children (9 to 11) for Flanders, Finland and Russia where this age category was to a large extent underrepresented.

Rule 3: Always use some of the filters. Always check which filter you are currently using.

- **Weights:** Weights should be used for descriptive analyses (frequencies, cross-tabulations). This dataset only includes country-related weights. Data from Finland, Flanders and Russia were not weighted. We do not have European weights, since we are not representative for Europe.

Rule 4: Use the variable 'weights' for descriptive analysis of all variables besides modules; use 'weights_modules' for analyses including modules.

All details are available in the 'Data Dictionary'. More details concerning sampling etc. are in the technical report. If needed, contact the data manager at eukodatamanager@gmail.com

Annex 3: Key variables

Below is a selected list of the key variables, including the source of the adapted variables (listed at the end). Full questionnaires are available at http://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online/toolkit/ .

Table 8: Child identity and resources

Concept	Variable	Questions/Response options	Variable labels	Reference
Gender	c_QA1	What would you say is your sex/gender?	A boy A girl I don't know Prefer not to say	
Age	c_QA2a	In what MONTH were you born?	January February March April May June July August September October November December I don't know Prefer not to say	
Age	c_QA2b	In what YEAR were you born?	1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009	
Living situation		*Thinking about the home where you live all or most of the time, do any of these people live there? If you live an equal amount of time on several places, please think about the home where you will be sleeping tonight.*	No Yes I don't know Prefer not to say	Currie, Griebler, Inchley, Theunissen, Molcho, Samdal, Dür (2010) Adapted from HBSC - Family Culture
	c_QA3a	Mother(s)		
	c_QA3b	Father(s)		
	c_QA3c	Stepfather/Partner of my mother		
	c_QA3d	Stepmother/Partner of my father		
	c_QA3e	Grandparent(s) or other relatives		
	c_QA3f	Siblings (including half, step or foster siblings)		
	c_QA3g	Other people		
	c_QA3h	I live alone		
	c_QA3i	I don't know		
	c_QA3j	Prefer not to say		

Table 9: Access and use

Concept	Variable	Questions/Response options	Variable labels	Reference
Devices used to go online	c_QB5a c_QB5b c_QB5c c_QB5h	*How often do you go online or use the internet using the following devices?* A mobile phone/smartphone A desktop computer, laptop or notebook computer A tablet Other	Never Hardly ever At least every month At least every week Daily or almost daily Several times each day Almost all the time I don't know Prefer not to say	
Time spent online	c_QB7 c_QB8	*About how long do you spend on the internet?* During a regular weekday (schoolday) During a regular weekend-day	Little or no time About half an hour About 1 hour About 2 hours About 3 hours About 4 hours About 5 hours About 6 hours About 7 hours or more I don't know Prefer not to say	

Table 10: Access and use

Concept	Variable	Questions/Response options	Variable labels	Reference
Online activities		*How often have you done these things ONLINE in the past month?*	Never Hardly ever	Helsper, van Deursen, Eynon, 2015
	c_QC3a	I looked for information about work or study opportunities	At least every week Daily or almost daily	
	c_QC3b	I used the internet for schoolwork	Several times each day	
	c_QC3c	I used the Internet to talk to people from other countries	Almost all the time I don't know Prefer not to say	
	c_QC3d	I looked for news online		
	c_QC3e	I got involved online in a campaign, protest or I signed a petition online		
	c_QC3f	I discussed political or social problems with other people online		
	c_QC3g	I created my own video or music and uploaded it to share		
	c_QC3h	I visited a social networking site		
	c_QC3i	I communicated with family or friends		
	c_QC3j	I played online games		
	c_QC3k	I watched video clips		
	c_QC3l	I listened to music online		
	c_QC3m	I participated in an online group where people share my interests or hobbies		
	c_QC3n	I looked for health information for myself or someone I know		
	c_QC3o	I browsed for things to buy or see what things cost		

Table 11: Risks

Concept	Variable	Questions/Response options	Variable labels	Reference
Overall negative online experiences	c_QF01	In the PAST YEAR, has anything EVER happened online that bothered or upset you in some way (e.g., made you feel upset, uncomfortable, scared or that you shouldn't have seen it)?	No Yes I don't know Prefer not to say	
Frequency of overall negative online experiences	c_QF02_rt	In the PAST YEAR, how often did this happen?	A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say	
Seeking support by child after negative online experience		*The last time something happened online that bothered or upset you, did you talk to anyone of these people about it?*	No Yes I don't know Prefer not to say	
	c_QF04a_rt	My mother or father (or step/foster mother or father)		
	c_QF04b_rt	My brother or sister (or step/foster/half sibling)		
	c_QF04c_rt	A friend around my age		
	c_QF04d_rt	A teacher		
	c_QF04e_rt	Someone whose job it is to help children		
	c_QF04f_rt	Another adult I trust		
	c_QF04g_rt	Someone else		
	c_QF04h_rt	I didn't talk to anyone		
	c_QF04i_rt	I don't know		
	c_QF04j_rt	Prefer not to say		
Reaction to overall negative online experiences		*The last time you had problems with something or someone online that bothered or upset you in some way, did you do any of these things afterwards?*	No Yes I don't know Prefer not to say	
	c_QF05a_rt	I ignored the problem or hoped the problem would go away by itself		
	c_QF05b_rt	I closed the window or app		
	c_QF05c_rt	I felt a bit guilty about what went wrong		
	c_QF05d_rt	I tried to get the other person to leave me alone		
	c_QF05e_rt	I tried to get back at the other person		
	c_QF05f_rt	I stopped using the internet for a while		
	c_QF05g_rt	I deleted any messages from the other person		
	c_QF05h_rt	I changed my privacy/contact settings		
	c_QF05i_rt	I blocked the person from contacting me		
	c_QF05j_rt	I reported the problem online (e.g., clicked on a 'report abuse' button, contacted an internet advisor or Internet Service Provider (ISP))		
	c_QF05k_rt	Something else		
	c_QF05l_rt	I don't know		
	c_QF05m_rt	Prefer not to say		

Table 12: Digital Ecology

Concept	Variable	Questions/Response options	Variable labels	Reference
Digital ecology		*How often does the following apply to you?*	Never Sometimes Often Always I don't know Prefer not to say	
	c_QD2a	I feel safe on the internet		
	c_QD2b	I find other people are kind and helpful on the internet		
Knowing what to do if someone acts online in a way children don't like	c_QD2c	I know what to do if someone acts online in a way I don't like	Never Sometimes Often Always I don't know Prefer not to say	
Tendencies in online self-disclosure	c_QD2d	I find it easier to be myself online than when I am with people face-to-face	No Yes I don't know Prefer not to say	Smahel, Brown, Blinka, 2012
	c_QD2e	I talk about different things online than I do when speaking to people face-to-face		
	c_QD2f	I talk about personal things online which I do not talk about with people face-to-face		

Table 13: Skills

Concept	Variable	Questions/Response options	Variable labels	Reference
Digital skills		*On a scale from 1 to 5 where 1 is 'Not at all true of me' and 5 is 'Very true of me', how true are these of you?*	Not true of me; Somewhat not true of me; Neither true nor not true of me; Somewhat true of me; Very true of me; I don't know; Prefer not to say	Helsper, van Deursen, Eynon, 2015
	c_QE1a_oy	I know how to save a photo that I find online		
	c_QE1b_oy	I know how to change my privacy settings (e.g., on a social networking site)		
	c_QE1c_oy	I find it easy to check if the information I find online is true		
	c_QE1d_oy	I find it easy to choose the best keywords for online searches		
	c_QE1e_oy	I know which information I should and shouldn't share online		
	c_QE1f_oy	I know how to remove people from my contact lists		
	c_QE1g_oy	I know how to create and post online video or music		
	c_QE1h_oy	I know how to edit or make basic changes to online content that others have created		
	c_QE1i_oy	I know how to install apps on a mobile device (e.g., phone or tablet)		
	c_QE1j_oy	I know how to keep track of the costs of mobile app use		
	c_QE1k_oy	I know how to make an in-app purchase		

Table 14: Risks

Concept	Variable	Questions/Response options	Variable labels	Reference
Contact with unknown people online		*In the PAST YEAR, how often have you done these things online?*	Never A few times	Livingstone, Helsper, 2007
	c_QF10a	Looked for new friends or contacts on the internet	At least every month	
	c_QF10b	Sent my personal information (e.g., my full name, address or phone number) to someone I have never met face-to-face	At least every week Daily or almost daily	
	c_QF10c	Added people to my friends or contacts I have never met face-to-face	I don't know Prefer not to say	
	c_QF10d	Pretended to be a different kind of person online from who I really am		
	c_QF10e	Sent a photo or video of myself to someone I have never met face-to-face		
Online interaction with unknown people	c_QF11	Have you EVER had contact on the internet with someone you have not met face-to-face before?	No Yes I don't know Prefer not to say	
Meeting unknown people offline	c_QF12_rt1	In the PAST YEAR, have you EVER met anyone face-to-face that you first got to know on the internet?	No Yes I don't know Prefer not to say	
Reactions to offline meeting	c_QF13_rt2	Thinking of the LAST TIME you met anyone face-to-face that you first got to know on the internet, how did you feel about it?	I was happy I was not happy or upset I was a little upset I was fairly upset I was very upset I don't know Prefer not to say	
Overall victimization (online or offline)	c_QF20	In the PAST YEAR, has anyone EVER treated you in such a hurtful or nasty way?	No Yes I don't know Prefer not to say	
Online and offline victimization		*In the PAST YEAR, how often did this happen in any of the following ways?*	Never A few times	
	c_QF21a_rt1	In person face-to-face (a person who is together with you in the same place at the same time)	At least every month	
	c_QF21b_rt1	Via a mobile phone or internet, computer, tablet, etc.	At least every week Daily or almost daily	
	c_QF21c_rt1	Some other way	I don't know Prefer not to say	
Types of online victimization		*Have any of these things happened to you in the last year?*	No Yes	
	c_QF23a_rt2	Nasty or hurtful messages were sent to me	I don't know Prefer not to say	
	c_QF23b_rt2	Nasty or hurtful messages were passed around or posted where others could see		
	c_QF23c_rt2	I was left out or excluded from a group or activity on the internet		
	c_QF23d_rt2	I was threatened on the internet		
	c_QF23e_rt2	I was forced to do something I did not want to do		
	c_QF23f_rt2	Other nasty or hurtful things happened to me on the internet		

Intensity of harm due online victimization	c_QF24_rt2	Thinking of the LAST TIME someone treated you in a hurtful or nasty way ONLINE, how did you feel?	I was not upset I was a little upset I was fairly upset I was very upset I don't know Prefer not to say
Overall aggression (online or offline)	c_QF28	In the PAST YEAR, have you EVER TREATED someone else in a hurtful or nasty way?	No Yes I don't know Prefer not to say
Online and offline aggression		*In the PAST YEAR, how often have you TREATED someone else in any of the following ways?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say
	c_QF29a_rt	In person face-to-face (a person who is together with you in the same place at the same time)	
	c_QF29b_rt	Via a mobile phone or internet, computer, tablet, etc.	
	c_QF29c_rt	Some other way	
Seeing sexual images	c_QF30	In the PAST YEAR, have you EVER SEEN any sexual images?	No Yes I don't know Prefer not to say
Seeing sexual images - platforms		*In the PAST YEAR, how often have you seen images of this kind in any of the following ways?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say
	c_QF31a_rt	In a magazine or book	
	c_QF31b_rt	On television, film	
	c_QF31c_rt	Via a mobile phone, computer, tablet or any other online device	
Reactions to seeing sexual images	c_QF32_rt	Thinking of the LAST TIME you have seen images of this kind, how did you feel about it?	I was happy I was not happy or upset I was a little upset I was fairly upset I was very upset I don't know Prefer not to say
Receiving sexual messages	c_QF40_oy	In the PAST YEAR, have you EVER RECEIVED any sexual messages? This could be words, pictures or videos?	No Yes I don't know Prefer not to say
Sending sexual messages	c_QF45_oy	In the PAST YEAR, have you EVER SENT or POSTED any sexual messages? This could be words, pictures or videos about you or someone else.	No Yes I don't know Prefer not to say
Frequency of sending sexual messages		*In the PAST YEAR, how often, if ever, have you SENT or POSTED any sexual MESSAGES (words, pictures or videos) in the following ways?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say
	c_QF46a_rt_oy	I have sent someone a sexual message (e.g., words, pictures or video)	
	c_QF46b_rt_oy	I have posted a sexual message (e.g., words, pictures or video) where other people could see it on the internet	
	c_QF46c_rt_oy	I have asked someone on the internet for sexual information about him or herself (like what his or her body looks like without clothes on or sexual things he or she has done)	

Unwanted request for sexual information	c_QF47_oy	In the PAST YEAR, how often, if ever, have you been asked by someone on the internet for sexual information (words, pictures or videos) about yourself (like what your body looks like without clothes on or sexual things you have done) when you did not want to answer such questions?	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say	Mitchell, Jones, 2012
Harmful user-generated content		*In the PAST YEAR, have you seen online content or online discussions where people talk about or show any of these things?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say	
	c_QF50a_oy	Ways of physically harming or hurting themselves		
	c_QF50b_oy	Ways of committing suicide		
	c_QF50c_oy	Ways to be very thin (such as being anorexic or bulimic, or "thinspiration")		
	c_QF50d_oy	Hate messages that attack certain groups or individuals (e.g., people of different colour, religion, nationality, or sexuality)		
	c_QF50e_oy	Their experiences of taking drugs		
	c_QF50f_oy	Gory or violent images, for example of people hurting other people or animals?		
Data misuse		*In the PAST YEAR, has any of the following happened to you on the internet?*	No Yes I don't know Prefer not to say	
	c_QF60a	Somebody used my personal information in a way I didn't like		
	c_QF60b	The device (e.g., phone, tablet, computer) I use got a virus or spyware		
	c_QF60c	I lost money by being cheated on the internet		
	c_QF60d	Somebody used my password to access my information or to pretend to be me		
	c_QF60e	Somebody created a page or image about me that was hostile or hurtful		
	c_QF60f	I spent too much money on in-app purchases or in online games		
	c_QF60g	Someone found out where I was because they tracked my phone or device		
Excessive internet use		*In the PAST YEAR, how often have these things happened to you?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say	Smahel, Blinka, 2012
	c_QF70a_oy	I have gone without eating or sleeping because of the internet		
	c_QF70b_oy	I have felt bothered when I cannot be on the internet		
	c_QF70c_oy	I have caught myself using the Internet although I'm not really interested		
	c_QF70d_oy	I have spent less time than I should with either family, friends or doing schoolwork because of the time I spent on the internet		
	c_QF70e_oy	I have tried unsuccessfully to spend less time on the internet		
	c_QF70f_oy	I have experienced conflicts with family or friends because of the time I spent on the internet		

Sharenting		*In the PAST YEAR, how often has this happened to you?*	Never A few times At least every month At least every week Daily or almost daily I don't know Prefer not to say
	c_QF80a_oy	My parent/carer published information (such as text, pictures or movies) about me on the internet without asking first if I was OK with it	
	c_QF80b_oy	I received negative or hurtful comments from someone because of something my parent/carer published online	
	c_QF80c_oy	I asked my parent/carer to remove things they had published on the internet	
	c_QF80d_oy	I was upset because of information my parents published online	

Table 15: Child identity and resources

Concept	Variable	Questions/Response options	Variable labels	Reference
Psychological Difficulties - Conduct Problems		*How true are these things of you?*	Not true for me A bit true for me Fairly true for me Very true for me I don't know Prefer not to say	Goodman, Meltzer, Bailey, 1998 Strenghts and Difficulties Q
	c_QA10a	I get very angry and often lose my temper		
	c_QA10d	I am often accused of lying or cheating		
	c_QA10e	I take things that are not mine from home, school or elsewhere		
Psychological Difficulties - Emotional Symptoms		*How true are these things of you?*	Not true for me A bit true for me Fairly true for me Very true for me I don't know Prefer not to say	Goodman, Meltzer, Bailey, 1998 Strenghts and Difficulties Q
	c_QA11a	I worry a lot		
	c_QA11b	I am nervous in certain new situations, I easily lose confidence		
	c_QA11d	I am often unhappy, sad or tearful		
	c_QA11e	I have many fears and I am easily scared		
Psychological Difficulties - Hyperactivity		*How true are these things of you?*	Not true for me A bit true for me Fairly true for me Very true for me I don't know Prefer not to say	Goodman, Meltzer, Bailey, 1998 Strenghts and Difficulties Q
	c_QA12a	I am restless, I cannot stay still for long		
	c_QA12d	I am easily distracted and find it difficult to concentrate		
	c_QA12e	I think before I do things		
Perceived discrimination		*Do you sometimes feel that you are treated badly because of the following?*	Never Hardly ever Sometimes Often Very often I don't know Prefer not to say	Williams, González, Williams, Mohammed, Moomal, Stein, 2008 Major Experiences of Discriminatio n (expanded version) adapted
	c_QA16a_oy	Because of where my family is from		
	c_QA16b_oy	Because of my skin colour		
	c_QA16c_oy	Because of my religion		
	c_QA16d_oy	Because of my height or weight		
	c_QA16e_oy	Because of a disability		
	c_QA16f_oy	Because of not having enough money		
	c_QA16g_oy	Because of the type of people I fall in love with		
	c_QA16h_oy	Because of how I look or behave		
	c_QA16i_oy	Because of my opinions or beliefs		
	c_QA16j_oy	For some other reason		
Sensation Seeking		*How true are these things of you?*	Not true A bit true Fairly true Very true I don't know Prefer not to say	Slater, 2003
	c_QA18a	I do dangerous things for fun		
	c_QA18b	I do exciting things, even if they are dangerous		
Self-efficacy		*How true are these things of you?*	Not true A bit true Fairly true Very true I don't know Prefer not to say	Schwarzer, Jerusalem, 1995
	c_QA21c	It's easy for me to stick to my aims and achieve my goals		
	c_QA21d	I am confident that I can deal with unexpected problems		
	c_QA21e	I can generally work out how to handle new situations		
	c_QA21f	I can solve most problems if I try hard		
	c_QA21i	If I am in trouble I can usually think of something to do		

Table 16: Well-being

Concept	Variable	Questions/Response options	Variable labels	Reference
Life Satisfaction	c_QH1	Here is a picture of a ladder. Imagine that the top of the ladder '10' is the best possible life for you and the bottom '0' is the worst possible life for you. In general, where on the ladder do you feel you stand at the moment? PLEASE TICK THE BOX NEXT TO THE NUMBER THAT BEST DESCRIBES WHERE YOU STAND.	10 - Best possible life 9 8 7 6 5 4 3 2 1 0 - Worst possible life I don't know Prefer not to say	Currie, Griebler, Inchley, Theunissen, Molcho, Samdal, Dür, 2010

Table 17: Family

Concept	Variable	Questions/Response options	Variable labels	Reference
Family environment		*How true are the following things about your family and home?*	Not true A bit true	WHO, 2016
	c_QI2a	When I speak someone listens to what I say	Fairly true Very true	Zimet, Dahlem,
	c_QI2b	My family really tries to help me	I don't know Prefer not to say	Zimet, Farley,
	c_QI2c	I feel safe at home		1988 EUKO IV original
Parenting		*How often do the following things apply to you? If you live an equal amount of time at several places, think about the home where you will be sleeping tonight.*	Never Hardly ever Sometimes	The ESPAD Group, 2016
	c_QI3a	My parent/carer praises me for behaving well	Often Very often	
	c_QI3b	My parent/carer sets rules about what I can do at home	I don't know Prefer not to say	
Parental active mediation		*When you use the internet, how often does your parent/carer do any of these things?*	Never Hardly ever Sometimes	
	c_QI4a	Encourages me to explore and learn things on the internet	Often Very often	
	c_QI4b	Suggests ways to use the internet safely	I don't know Prefer not to say	
	c_QI4c	Talks to me about what I do on the internet		
	c_QI4j	Helps me when something bothers me on the internet		
Seeking support related to internet use		*Have you EVER done any of these things?*	Never Hardly ever Sometimes	
	c_QI5a	Told my parent/carer about things that bother or upset me on the internet	Often Very often	
	c_QI5b	Helped my parent/carer to do something they found difficult on the internet	I don't know Prefer not to say	
	c_QI5f	Asked for my parent's/carer's help with a situation on the internet that I could not handle		
Restrictions from parents		*Does your parent/carer allow you to do the following things on the internet and if so, do you need their permission to do them?*	I am allowed to do this anytime; I am allowed to do this only with permission or supervision; I am not allowed to do this; I do not know if I am allowed to do this; I don't know Prefer not to say	
	c_QI6a	Use a web or phone camera (e.g., for Skype or video chat)		
	c_QI6b	Download music or films		
	c_QI6c	Use a social networking site (e.g., Facebook,Snapchat, Instagram, Twitter)		
Technical parental mediation		*Does your parent/carer make use of any of the following…?*	No Yes I don't know Prefer not to say	
	c_QI7a	Parental controls or other means of blocking or filtering some types of content		
	c_QI7b	Parental controls or other means of keeping track of the Internet content I look at or apps I use		
	c_QI7i	Technology to track where I am (such as GPS)		
Ignoring parental mediation	c_QI13	Do you ever ignore what your parent/carer tells you about how and when you can use the internet?	No Yes, sometimes Yes, often I don't know Prefer not to say	

Table 18: School

Concept	Variable	Questions/Response options	Variable labels	Reference
School environment		*Here are some statements about your school and the students and teachers in your school. Please say how much you agree or disagree with each one.*	Not true A bit true Fairly true	Currie, Griebler, Inchley,
	c_QJ1a	I feel like I belong in my school	Very true I don't know	Theunissen, Molcho,
	c_QJ1b	I feel safe at school	Prefer not to say	Samdal, Dür, 2010
	c_QJ1c	Other students are kind and helpful		HBSC -
	c_QJ1d	Teachers care about me as a person		adapted
	c_QJ1e	There is at least one teacher I can go to if I have a problem		
Teacher mediation		*Have any teachers at your school done these things?*	Never	
	c_QJ2a	Suggested ways to use the internet safely	Hardly ever Sometimes	
	c_QJ2b	Encouraged me to explore and learn things on the internet	Often Very often	
	c_QJ2c	Made rules about what I can do on the internet at school	I don't know Prefer not to say	
	c_QJ2h	Helped me in the past when something has bothered me on the internet		

Table 19: Peers and community

Concept	Variable	Questions/Response options	Variable labels	Reference
Social support from friends		*How true are the following things for you?*	Not true A bit true	Zimet, Dahlem,
	c_QK1a	My friends really try to help me	Fairly true	Zimet,
	c_QK1b	I can count on my friends when things go wrong	Very true I don't know	Farley, 1988
	c_QK1c	I can talk about my problems with my friends	Prefer not to say	
Active mediation by friends		*Have any of your friends done these things?*	Never	
	c_QK2a	Suggested ways to use the internet safely	Hardly ever Sometimes	
	c_QK2b	Encouraged me to explore and learn things on the internet	Often Very often I don't know Prefer not to say	

Sources of adapted measures:

- Currie, C., Griebler, R., Inchley, J., Theunissen, A., Molcho, M., Samdal, O., & Dür, W. (2010). *Health Behaviour in School-aged Children (HBSC) study protocol: Background, methodology and mandatory items for the 2009/10 survey.* Edinburgh: CAHRU & Vienna: LBIHPR.
- EUKO (EU Kids Online) (2014). Research toolkit. Available at lse.ac.uk/EUKidsOnline/Toolkit
- Goodman R., Meltzer H., Bailey V. (1998) The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry, 7*, 125-130. doi:10.1007/s007870050057
- Goodman, E., Adler, N. E., Kawachi, I., Frazier, A. L., Huang, B., & Colditz, G. A. (2001). Adolescents' perceptions of social status: Development and evaluation of a new indicator. *Pediatrics, 108*(2), e31-e31.
- Helsper, E.J., van Deursen, A.J.A.M. and Eynon, R. (2015). From digital skills to tangible outcomes: Full questionnaire. Available at www.lse.ac.uk/media@lse/research/Research-Projects/From-Digital-Skills-to-Tangible-Outcomes/Pdf/From-Digital-Skills-to-Tangible-Outcomes-Questionnaire.pdf
- Livingstone, S., and Helsper, E. J. (2007). Taking risks when communicating on the internet: The role of offline social-psychological factors in young people's vulnerability to online risks. *Information, Communication and Society, 10*(5), 619-643.
- Mitchell, K.J. & Jones, L.M. (2012). *Youth Internet Safety Study: Methodology Report.* Durham, NH: Crimes against Children Research Center. Available at https://scholars.unh.edu/cgi/viewcontent.cgi?article=1049&context=ccrc
- Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio, 1*, 35–37. Windsor, UK: NFER-NELSON. Retrieved from https://www.researchgate.net/publication/284672098_Measures_in_Health_Psychology_A_User's_Portfolio_Causal_and_Control_Beliefs
- Slater, M. D. (2003). Alienation, Aggression, and Sensation Seeking as Predictors of Adolescent Use of Violent Film, Computer, and Website Content. *Journal of Communication, 53*(1), 105–121. https://doi.org/10.1093/joc/53.1.105
- Smahel, D., & Blinka, L. (2012). *Excessive internet use among European children*. In S. Livingstone, L. Haddon, & A. Görzig (Eds.), Children, risk and safty on the internet, 191-203. Bristol: The Policy Press.
- Smahel, D., Brown, B. B., & Blinka, L. (2012). Associations between online friendship and Internet addiction among adolescents and emerging adults. *Developmental psychology*, *48*(2), 381-388.
- The ESPAD Group. (2016). *ESPAD Report 2015: Results from the European School Survey Project on Alcohon and Other Drugs.* Luxembourg: Publications Office of the European Union. Available at http://www.espad.org/sites/espad.org/files/ESPAD_report_2015.pdf
- WHO. (2016). Growing up unequal: gender and socioeconomic differences in young people' s health and well-being. Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey. (J. C. Inchley, D. B. Currie, T. Young, O. Samdal, T. Torsheim, L. Augustson, … V. Barnekow, Eds.), *Health Policy for Children and Adolescents, 7*. WHO Regional Office for Europe.
- Williams, D.R., González, H.M., Williams, S., Mohammed, S.A., Moomal, H, and Stein, D.J. (2008). Perceived Discrimination, Race and Health in South Africa: Findings from the South Africa Stress and Health Study. *Social Science and Medicine, 67*, 441-452.
- Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, *52*(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2

eukidsonline.net

Contact
Rostislav Zlamal and Hana Machackova
Interdisciplinary Research Team on Internet and Society
Faculty of Social Studies
Masaryk University
Jostova 10, 602 00 Brno
Czech Republic
E-mail: eukodatamanager@gmail.com