

YouTH TALKS



RESEARCH KIT

Oct 2023

HIGHER EDUCATION
FOR GOOD
TOWARD SUSTAINABLE FUTURES



THE YOUTH TALKS INITIATIVE

Youth Talks is the **largest-ever global consultation** inviting people aged **15-29** to share their opinions, concerns and aspirations for the future.

The project uses **state-of-the-art A.I.** technology to analyze responses, enabling participants to answer **open-ended questions** in thoughtful, nuanced ways and distilling vast amounts of shared content into rich, comprehensible and actionable insights.

Some figures from the first edition :

.1 000 000 contributions (answers to open or closed-ended questions)

.45 000 participants (people who responded to at least one question)

.212 countries and territories

.55 partners, including the UN-supported Principles for Responsible Management Education (PRME), the Club of Rome, top universities and world youth organizations.

<https://youth-talks.org/en/>

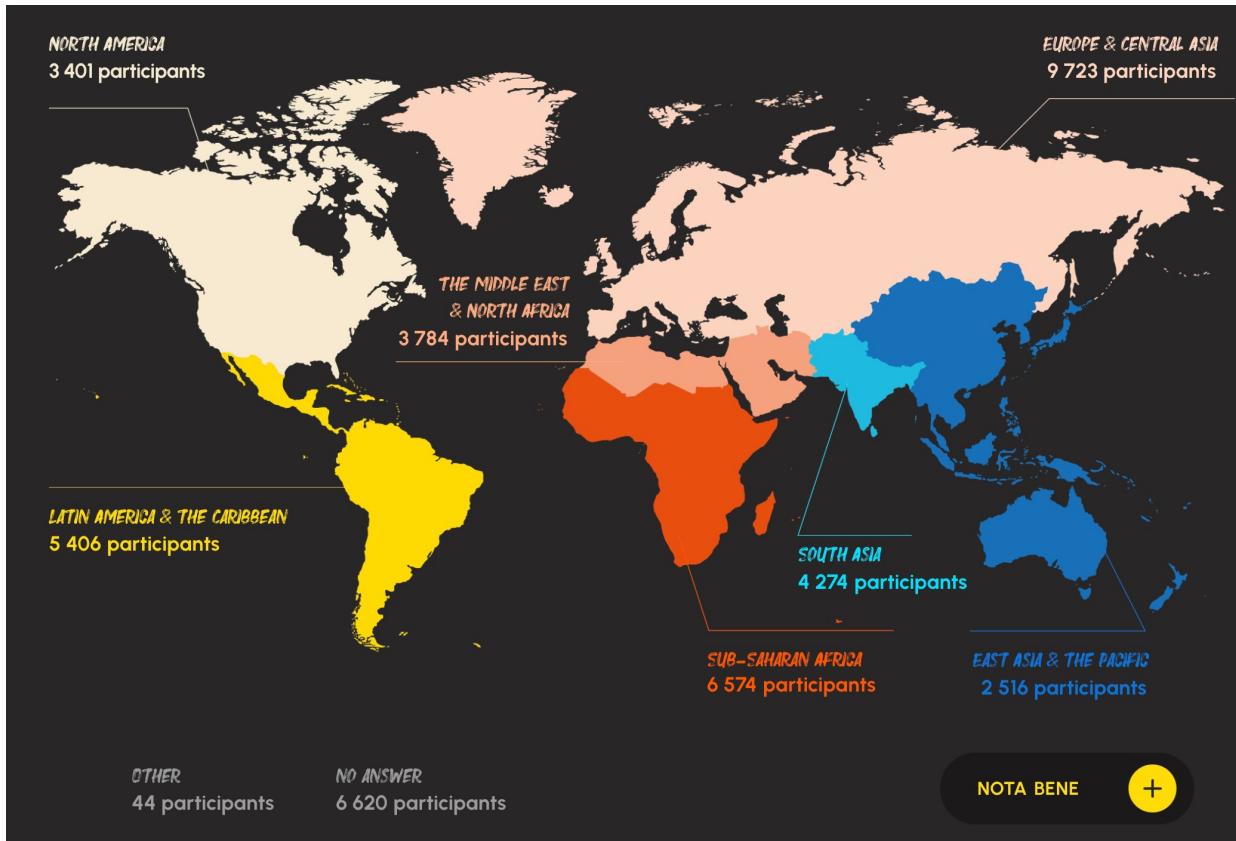


YOUTH
TALKS

QUESTIONS ASKED

Open-Ended Questions	Segmentation Questions	Empathy Questions (Research validated questionnaire from a 16-item version of Davis (1980, 1983) Interpersonal Reactivity Index *)
<p>1. When I think about the future, what I wish... for myself</p> <p>2. When I think about the future, what I wish... for the world (others, society, planet...)</p> <p>3. When I think about the future, what worries me... for myself</p> <p>4. When I think about the future, what worries me... for the world (others, society, planet...)</p> <p>5. What collective issues do we need to address to build the future I want?</p> <p>6. To build this desired future, I would be ready to give up:</p> <p>7. On the contrary, I would not be willing to give up (habits, lifestyles, values...):</p> <p>8. Why?</p> <p>9. To build this desired future, what we must all learn (knowledge, skills, behaviors)... at school is:</p> <p>10. To build this desired future, what we must all learn (knowledge, skills, behaviors)... in life is:</p> <p>11. What question would you like to ask young people around the world ?</p>	<p>1. My current situation</p> <p>2. My year of birth</p> <p>3. My gender</p> <p>4. My level of education (completed or in progress)</p> <p>5. My parents' highest level of education</p> <p>6. Here is an income scale in which 1 indicates the lowest income group and 10 indicates the highest income group in your country. In which group do you consider your family to be?</p>	<p>1. I often have tender, concerned feelings for people less fortunate than me.</p> <p>2. I really get involved with the feelings of the characters in a novel.</p> <p>3. In emergency situations, I feel apprehensive and ill-at-ease.</p> <p>4. I try to look at everybody's side of a disagreement before I make a decision.</p> <p>5. When I see someone being taken advantage of, I feel kind of protective toward them.</p> <p>6. I sometimes try to understand my friends better by imagining how things look from their perspective.</p> <p>7. After seeing a play or movie, I have felt as though I were one of the characters.</p> <p>8. Being in a tense emotional situation scares me.</p> <p>9. When I see someone being treated unfairly, I sometimes don't feel very much pity for them.</p> <p>10. I would describe myself as a pretty soft-hearted person.</p> <p>11. When I watch a good movie, I can very easily put myself in the place of a leading character.</p> <p>12. I tend to lose control during emergencies.</p> <p>13. When I'm upset at someone, I usually try to "put myself in his shoes" for a while.</p> <p>14. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.</p> <p>15. When I see someone who badly needs help in an emergency, I go to pieces.</p> <p>16. Before criticizing somebody, I try to imagine how I would feel if I were in their place.</p>

PARTICIPANTS MAP



NB: To make analysis and comparison easier, we have grouped countries together according to the nomenclature of the different regions used by the World Bank. However, not all regions have the same weight in the overall data as others and, within some regions, certain countries and sub-regions have varying levels of representation. This is particularly the case for Europe and Central Asia, in which Western Europe, Northern Europe, and Southern Europe are over-represented (9,313 participants) in comparison with Eastern Europe and Central Asia (348 participants). We find the same situation in East Asia and Pacific, with the countries of East Asia (particularly China) featuring much more strongly than Oceania.

MATERIAL AVAILABLE TO RESEARCHERS (1/3)

1. Main Database

The Youth Talks database is the most comprehensive youth-centred dataset ever compiled. This resource categorises written responses both by individual participant and by specific questions asked.

After cleaning, the final Youth Talks **dataset contains 41,753 records for as many participants**, and includes

- 40,999 participants with at least one response to the **segmentation questions**,
- 28,846 participants with at least one response to the **open-ended questions**,
- 15,874 participants with at least one response to the **empathy questions**.

For more information on the dataset, please request the **Data Quality Assessment Report**.

A detailed breakdown of **the distribution of participants across different segments** can be found on the Global Report on Youth Talks online platform.

A sample dataset is included in the research kit or can be provided on request.

MATERIAL AVAILABLE TO RESEARCHERS (2/3)

2. Sound Library & Image Library

Throughout the consultation phase, participants were granted the flexibility to express themselves not only through written words but also via voice recordings or visual illustrations.

As a result, sound and image repositories have been curated for every open-ended question, with the exceptions of questions 8 and 11. In total, **our collection encompasses 1,374 audio recordings and 764 images**. These assets are available upon request.

3. Video Interviews Library

The "Youth Talks" initiative further enriches its resources with a collection of **80 video interviews**, all of which can be found on the official "Youth Talks" YouTube Channel.

Interviewees hail from a diverse set of nations including France, the Philippines, Romania, Finland, China, Ivory Coast, South Africa, Lebanon, Brazil, Morocco, the USA, and Australia. This selection was intentional to ensure representation from every significant global region. Within these nations, interviewees were meticulously chosen to encapsulate various socio-economic backgrounds. In addition to the video footage, **transcripts of all interviews are available**, further serving as a valuable resource for researchers.

MATERIAL AVAILABLE TO RESEARCHERS (3/3)

4. Empathy Questionnaire

In addition to the Youth Talks consultation based on open-ended questions, we also provided a questionnaire on empathy to be completed by anyone who wanted to take things a step further. Empathy is a fundamental need in our interconnected society, facilitating mutual understanding and supporting conflict resolution. Faced with global challenges like climate change, social inequalities, and political polarization, empathy can elicit a sense of shared responsibility.

We chose to use a validated research questionnaire, the “Interpersonal Reactivity Index,” in order to guarantee the rigor of our survey.** This standardized tool means that our results can be compared with those of other studies, making the results easier to interpret as their dimensions are well defined and encouraging other researchers to use our data for their own quantitative work.

You can find a Case study using data collected via the empathy questionnaire of Youth Talks in the Research Kit.

** Ingoglia, S., Lo Coco, A., & Albiero, P. (2016). Development of a brief form of the Interpersonal Reactivity Index (B-IRI). *Journal of Personality Assessment*, 98(5), 461–471. <https://pubmed.ncbi.nlm.nih.gov/27050826/>

ACCESS PROCEDURES FOR THE YOUTH TALKS DATASET

As it stands, **Youth Talks dataset is not available for open access due to the sensitive nature of the information it contains and to certain agreements with our privileged partners.**

Access to this dataset is granted on a case-by-case basis and requires a formal request. Interested researchers are invited to apply through the 'Higher Education for Good Foundation' or via direct communication with any of our esteemed partner institutions.

Furthermore, for those preparing to submit a request, **a detailed Research Proposal template** can be found within the **Research Kit**. This tool is designed to streamline the application process and ensure that all necessary information is provided. If you do not have the Research Kit, it can be requested along with additional guidance on the application procedure.

Please find the contact information appended at the conclusion of this document for any inquiries or requests for the Research Kit. Your commitment to **responsible research** and adherence to the access protocol is greatly appreciated.

OTHER INFORMATION (1/4)

Scientific Process

Throughout the consultation —before, during, and after— collaborative efforts were undertaken by the Youth Talks Scientific Committee and the consultants from bluenove working on the project. Notably, the Committee was spearheaded by the outgoing Minister of Higher Education from France. Their joint endeavours were aimed at ensuring the data collection methodology's utmost rigor and reliability. Comprehensive details regarding this methodology are accessible within the **Methodology Note** hosted on the Youth Talks platform and available in the Research Kit. Given the absence of any directly analogous methodologies, the Committee leveraged established qualitative and quantitative methods while innovating a protocol specifically tailored to this initiative.

[Delve deeper into the Youth Talks Methodology Note](#)

Languages

The Youth Talks initiative facilitated consultations across **six languages: English, French, Chinese, Portuguese, Spanish, and Arabic.**

Interestingly, some participants ventured beyond their native tongue to respond. For the purposes of NLP-driven clustering analysis, **English served as the pivotal language**, meaning all submissions were translated into English prior to NLP analysis. Contributions in their original language are preserved within the database and available upon request.



OTHER INFORMATION (2/4)

Representativeness

A sample is said to be representative if it proportionally reflects the attributes of a population (e.g., gender, age, location, education, income).

Youth Talks (YT) is a convenience sample: participants have been included based on their accessibility and their willingness to participate. It may therefore not be fully representative of the age group of the population 15 to 29 years old.

To assess this possibility, we compare in [Figure 1](#) below the share of (possible to identify) 15-24 years old people who identify the most with "being a part of the world" in YT with the counterpart share in the probability-based and nationally representative samples provided by the [Changing Childhood Project](#) (CCP) for 21 countries.

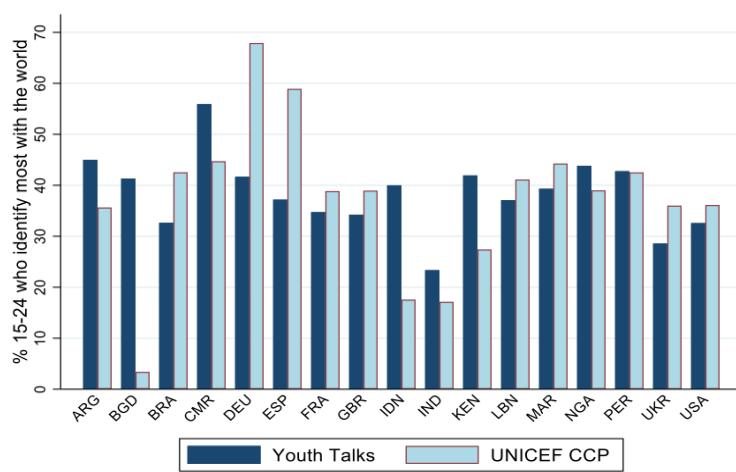
For most countries, the results are very similar, keeping in mind that the margin of error is +/- 4% (at the 95% confidence level) for CCP answers.

Overall, these results suggest that YT is a valuable tool not only to estimate treatment effects but also to get an idea of the prevalence of an opinion in a given population.

[Delve deeper into the Youth Talks Methodology Note](#)

Which of the following do you identify with MOST?

- Being part of the city or area where you live
- Being a part of this country
- Being a part of the world
- Don't know



Notes: at least 10 YT respondents; YT: simple average; UNICEF: representative sample

(Generalisation from sample to population could still be achieved ex post, for example through [multilevel regression with poststratification](#). In addition, when the interest is in a treatment effect (e.g., the effect of an attribute on social and political attitudes), online convenience samples [often provide similar results](#) to those obtained from nationally representative samples).

OTHER INFORMATION (3/4)



Clustering

Please note that, **80% of the responses have been systematically grouped into clusters and meta-clusters for ease of analysis. 15% to 20% of the verbatims are not linked with any clusters.** For each question, we have clustered 80 to 85% of the data during the analysis phase: as per our experience and methodology, the percentages of clusters do not vary significantly (less than 1%) above 75 to 80% of clusterised data. The time required to analyse the remaining amount, however, starts to increase a lot with the last verbatims. Stopping at 80% allows us to ensure the best quality for our deliverables without deteriorating the quality of our clusters. Therefore, percentages are calculated within the clusterised data but can be extrapolated for the entire dataset. To ensure full transparency, we always include verbatims with no clusters in the dataset.

Nuggets

Leveraging advanced NLP techniques, we were able to distill over 1,000 insightful "nuggets" (averaging roughly 100 per question) based on a spectrum of selection variables. These nuggets embody nuanced, cogently articulated, and unconventional viewpoints. You can find most of these Nuggets in the Question Boxes under the section: "Some remarkable answers..." . Access to the full dataset of nuggets can be facilitated upon request.



OTHER INFORMATION (4/4)

Format of the Dataset

The dataset is provisioned in three distinct formats to cater to diverse analytical requirements:

Tabular Formats: Encompasses a structured compilation of 127 columns and 41,753 records. This format is concise, containing up to three responses to open-ended questions for streamlined analysis.

CSV (Comma-Separated Values): Ideal for compatibility with a wide range of data analysis tools and platforms.

Excel: Offers a user-friendly interface for those accustomed to Microsoft Excel's features for data manipulation.

Comprehensive Format:

JSON (JavaScript Object Notation): This format encapsulates the entirety of responses to the open-ended questions, ensuring no loss of detail and facilitating complex data structuring needs.

Supplementary Excel File: This auxiliary file is designed to provide additional context and metadata regarding the dataset, which includes:

A supplementary Excel file:

- Definitions and details of each column in the tabular dataset.
- JSON Schema: A blueprint of the JSON file structure for validation and understanding of data organization.
- Macro-clusters and Clusters: Insightful categorization aiding in higher-level analysis and data segmentation.



YOUTH
TALKS

Contact



Marine.hadengue@higheredforgood.org