

TeensLab dataset: Economic preferences and cognitive abilities of 4,844 teenagers *

Mónica Vasco^{†‡}, Antonio Alfonso[†], Pablo Brañas-Garza[†],
Antonio Cabrales[§], Teresa García[¶], Diego Jorrat[†], Jaromír Kovářik^{||},
Pablo Lomas[†], María del Pino Ramos-Sosa[†], Patricia Román[†],
Ángel Sánchez[§], María José Vázquez-De Francisco^{†**}

November 28, 2023

Abstract

This project provides a database of teenagers from 33 different educational centers in Spain. We elicited 7,073 observations of students from 10 to 23 years old, in different educational stages (elementary, middle school, high school, and vocational training). Both the collection process and the available variables on economic preferences and abilities are described.

Keywords: Teenagers, lab-in-the-field experiment, economic preferences

JEL codes: C81, C91, D90

*This research was supported by the Spanish Ministry of Economy and Competitiveness (PID2021-126892NB-I00), Excelencia-Junta (PY-18-FR-0007) and Agencia Andaluza de Cooperación Internacional para el Desarrollo (AACID-0I008/2020).

[†]Corresponding author. Email: movaru09@gmail.com Address: 4 Escritor Castilla Aguayo Street (14004) Córdoba, Spain

[‡]*LoyolaBehLab*, Universidad Loyola Andalucía, Spain

[§]Universidad Carlos III, Spain

[¶]Universidad de Granada, Spain

^{||}Universidad del País Vasco, Spain

^{**}ETEA Foundation-Development Institute, Universidad Loyola Andalucía, Spain.

1 Introduction

It is well known that individual preferences and cognitive abilities are important determinants of decision-making, not only in modeling economic events but also in empirical studies (Dohmen et al., 2010; Eckel et al., 2005; Guiso and Paiella, 2004). For this reason, there is a great interest in understanding the outcomes of the decision-making behavior of individuals and the underlying motivations of the subjects in the design of public policies (Chapman, 2003).

In order to understand and predict economic behavior, it is essential to study individual attitudes towards risk preferences, since they are involved in most of the decisions that individuals must make throughout their lives (Sutter et al., 2013). Not only in the short term but also in the long term, for example, it has been found to be related to entering the labor market or earning higher salaries (Heckman et al., 2006). That is why there is a current awareness about conducting research with adolescents, as it is a key stage in their development (List et al., 2023). In this line, Golsteyn et al. (2014) explore how time preferences are related to social and economic outcomes across the adolescent lifetime (see Bettinger and Slonim (2007) and Castillo et al. (2011)). And also for risk taking decisions (Belzil and Leonardi, 2013; Dohmen et al., 2010; Dohmen et al., 2011; O'Donoghue and Rabin, 2001).

However, the implications are not only associated with economic results but go beyond that. Relevant research shows the relationship between time preferences and specific health problems such as smoking or obesity (Lawless et al., 2013). Thus, much of the research focuses on discussing aspects that predict or relate to adolescent risk behavior (Anderson and Mellor, 2008). Furthermore, O'Donoghue and Rabin (2001) observed that as discount rates decreased with age, adolescents tend to show little regard for future health consequences. Hence, programs that aim to raise the awareness of young people regarding healthier choices focus on showing the short-term consequences (e.g. in sexual relationships and behavior (Chesson et al., 2006)).

Also in the educational context, there are studies that relate certain behaviors to educational expectations, such as more patient adolescents believe that they are more likely to go to university (Anders and Micklewright, 2015; Belzil and Leonardi, 2013; Hanushek et al., 2020).

The policy implications therefore have scope for future economic outcomes for adolescents; if it is possible to design interventions to enhance cognitive abilities to lead the increase of patience and to acquire a higher willingness to take risks, this is an extraordinary additional effect that should definitely be applied (Dohmen et al., 2010).

In addition to the administrative data recorded, a few databases with more than 1000 observations are publicly available. [\[CITAR PAPERS CON DATASETS\]](#)

Our project aims to contribute to providing a dataset for future research in adolescents. We conducted our experiment (Lab-in-the-field) in 33 different educational centers. The experimental design allowed us to elicit 7,073 observations of Spanish students. The data contain unique identifiers and other general information in the first place. The complete experiment has several sets of variables: (1) sociodemographic survey, (2) time preferences elicitation, (3) risk preferences elicitation, (4) social preferences task, (5) Probabilistic beliefs tasks, (6) Abilities tasks, (7) Networks, (8) Bullying, (9) Strategical thinking, (10) Creativity.

Recent developments in the open data initiative have generated a growing interest of researchers in sharing datasets. This drive to use Information and Communication Technologies (ICT) to create open data initiatives for the purpose of advancing scientific research has gained visibility. The emergence of Web 2.0, which promotes peer-to-peer collaboration, interactivity and user-generated innovation, has further catalyzed the development of open data initiatives aimed at sharing and distributing information among diverse stakeholders to address societal problems Mergel et al., 2009.

The main benefits of sharing research datasets are manifold. First, it facilitates the ability to enrich and promote the progress of scientific research. The availability of open data allows researchers to build on previous datasets, encouraging them to generate new knowledge and foster advanced discoveries from old datasets Kaye et al., 2009. In addition, data sharing encourages researchers to reconsider and reinterpret the meaning of previous datasets in the light of modern thinking Bryn, 2009. It could be argued that sharing datasets by reconstructing and/or combining multiple existing datasets represents the essence and basis of knowledge generation in today's research. In short, the open data initiative has become an essential driver for the advancement of science and knowledge generation.

The structure of this article is straightforward. Section 2 describes the process and characteristics of how the data were collected and Section 3 describes all the available variables grouped by topic. Finally, Section 4 concludes with some endnotes.

2 Methodology

Other studies have collected data in schools. When dealing with minors, there are special protections and laws that must be rigorously followed. Our project was approved by the Ethical Committee of Universidad Loyola Andalucía. In addition, all responses are guaranteed to be anonymized and researchers are not able to recognize or associate subjects.

There are some additional complications associated with non-standard samples. To avoid missing data, the format of the answers is simplified, most of the questions are presented with multiple-choice answers instead of open answers. Also, subjects could not skip questions without answering due to the format of the software we used. However, for some questions that might be sensitive for subjects, we gave them the possibility to answer with an option of *"I would prefer not to specify"*.

We use STATA for the data pre-processing, except for the networks, for which we use R to obtain the measurements which are later incorporated into the complete database.

The dataset was posted on XXXX, it was published in different formats (xls, cvs, dta). We also included the STATA scripts for some basic summaries or analyses mentioned throughout the paper.

2.1 Data collection

The data sample was obtained through agreements with school directors who agreed to integrate the experiment into their pedagogical curriculum and to carry it out as a classroom activity. Consequently, we achieved an increase in participation, as predicted in Alfonso et al. (2023). We proceeded directly to the schools to run the experiment in an online format through a platform called SAND, which allows greater control over the privacy of the data. Students read the instructions and navigated through the questionnaire, which contained multiple screens, via their devices.

The entire instrument was administered in Spanish. All responses remained anonymous and subjects were paid with hypothetical payments since schools did not accept real money experiments. It has been demonstrated that teenagers do not change their behavior and decisions in the presence of hypothetical incentives in risk and time preferences Alfonso et al., 2023.

[CITAR PAPERS CON HYPOTHETICAL INCENTIVES]

3 Variables

The variables are grouped by topics. In bold we emphasize the name of the variable in the data set.

3.1 Sample variables

A total of 7,073 observations, sample variables include a unique identifier for each subject (***usuario_id***), as well as an identifier of the network to which the subject belongs (***net_id***). The ***yeardone*** variable indicates the year in which the experiment was run, 2021 to 2023. The ***privacy*** variable specifies when the subject has accepted the data protection before starting the experiment and it allows us to record the answers of the subject. We find that 16.73% of the subjects do not agree to do the experiment and therefore do not start the study. From the remainder who started the experiment, we found a drop-out rate of 10.34% (***dropout***).

In addition, another important issue in our study is that it has evolved over time, according to some of the results we observed. In other words, tasks have been included, withdrawn, or transformed over the period. However, most of the main variables remain the same and are comparable across the whole sample. Therefore, we have to take into account that for some specific tasks we only have a small subset of observations. Table X shows the changes in the experiment and the data available for each study identifier in each school (***study***).

[Incluir tabla con cambios de la encuesta a lo largo del tiempo]

3.2 School variables

Available information is available for each center where data is collected. The unique identifier for each high school is *school*. Although most of the high schools are public, there are semi-public ones as well (*schooltype*: 1 = Public, 2 = Semi-public). The sample was collected in two regions of Spain, 68.13% in 5 different provinces of Andalusia (11 different towns in total) and the remaining 31.87% in 8 different locations in Barcelona. Table 1 shows that information corresponds to the variables *schoolprov* (Province of the school) and *schooltown* (Location of the school), respectively.

Table 1: Distribution of observations by province and high school location

Town where the school is	School province						Total
	1	2	3	4	5	6	
1	1,112	0	0	0	0	0	1,112
2	193	0	0	0	0	0	193
3	245	0	0	0	0	0	245
4	0	501	0	0	0	0	501
5	0	105	0	0	0	0	105
6	0	908	0	0	0	0	908
7	0	457	0	0	0	0	457
8	0	0	170	0	0	0	170
9	0	0	0	475	0	0	475
10	0	0	0	0	483	0	483
11	170	0	0	0	0	0	170
12	0	0	0	0	0	1,249	1,249
13	0	0	0	0	0	55	55
14	0	0	0	0	0	422	422
15	0	0	0	0	0	160	160
16	0	0	0	0	0	163	163
17	0	0	0	0	0	58	58
18	0	0	0	0	0	48	48
19	0	0	0	0	0	99	99
Total	1,720	1,971	170	475	483	2,254	7,073

In addition, for each student, the *stage*, the year or *grade* (see Table 2), and also the *group* in which they are in are specified. In total, we get 255 classes in the whole sample. The size of each group can also be obtained with the variable *class_size*.

Table 2: Distribution of observations by stage and grade

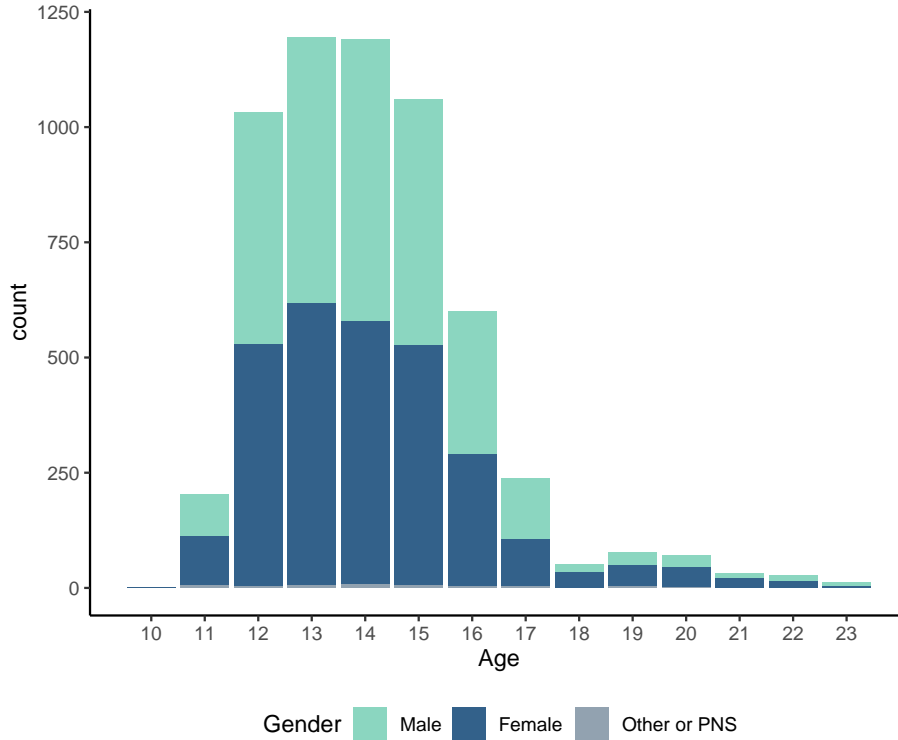
Stage	Year/Grade						Total
	1	2	3	4	5	6	
Elementary school	0	0	0	0	219	316	535
Middle school	1,676	1,733	1,461	1,178	0	0	6,048
High school	122	0	0	0	0	0	122
Vocational training	208	160	0	0	0	0	368
Total	2,006	1,893	1,461	1,178	219	316	7,073

3.3 Sociodemographic and related variables of individuals

Regarding individual-level attributes, the gender of the participants is coded with the variable *female* (0 = Male, 1 = Female). Actually, 47.38% of the sample is female and 48.62% male, the remaining part is unknown¹ (*gender* = 99). Their *age* is given in years but also the year of their birth is known (*yearbirth*).

¹Either because they did not want to answer or because they selected another category, or because they dropped out.

Figure 1: Sample distribution by age and gender



Family. On the other hand, information concerning their family is also included. Siblings and how many of them they have are reported, *siblings* and *nsiblings*², respectively. For those who report siblings, we also ask about their position among their siblings with *rankbrother*. The immigrant origin is also collected with the variable *migrant*. They can report that they or their parents were not born in Spain. Indeed our sample shows 14% of cases with an immigrant origin.

We also cover information on household income. However, aware of the difficulty of such a question for the subjects of this study, we decided to adapt it to their level. We asked it in a relative way: Imagine a stairway with ten(five) steps, where on step 1 are the poorest families in Spain, and on step 10 (5) are the richest ones. On which step of the stairway would you place your family? First, we use a 10-step version (*stairs10*) and then a 5-step version (*stairs5*), both of which are normalized in the *stairsN* variable.

GPA. We also asked them about their academic results, if they have obtained any A (*mark10*) or B (*mark8*), and how many of them in core subjects such as Maths, Literature, and English, but we also include an open category under "other subject". (*nmark10* and *nmark8*). The original answers with the courses are kept with the name *nmark10_text* and *nmark8_text*. We define a total variable where only the three main subjects are taken into account (*gpa*). this variable takes values from 0 to 9 according to the combination of A and B, as can be seen in Table 3. It is also normalized from 0 to 1 (*gpaN*).

Table 3: GPA score

		Num. A			
		0	1	2	3
Num. B	0	0	2	5	9
	1	1	4	8	x
	2	3	7	x	x
	3	6	x	x	x

Mood. Moreover, we included three questions about their mood regarding school and social relationships. We used a frequency Linkert scale for the three questions: Never/almost

²Note that in the latest surveys there is also a sister (*nsister*) and brother (*nbrother*) split variables.

never/sometimes/almost always/always

Q1 Over the last week, have you been doing well at high school? (*general*)

Q2 Over the last week, have you had fun with your friends? (*fun*)

Q3 Over the last week, have you felt lonely? (*alone*)

We created a total score from the *moodgeneral*, *moodfun* and *moodalone* variables, which include responses coded on a scale of 0-4 (0 = Never, 1 = Almost never, 2 = Sometimes, 3 = Almost always, 4 = Always). Note that the last question is on an inverted scale in the original. Then, we obtain *happy* variable with the total mood score (0-12). Also, *happyN* is presented with normalized values between 0 and 1.

Physical attributes. The *height_original* and *weight_original* variables contain the original raw responses, while the *height* and *weight* variables include information in centimeters and kilograms respectively. Although we do not have precise BMI information due to the fact that many subjects did not want to answer³ the height and weight questions, we include a task that requires self-identification using body shapes. A screen with 8 figures in a fixed order (from thinnest to fattest) is presented to the students. Depending on the gender they have indicated beforehand, male or female figures are presented. Therefore, *obesity* variable takes values from 1 to 8. However, there is also the possibility of not responding to this question (*obesity* = 99).

3.4 Time preferences

In order to elicit temporal preferences we use *The truck task*, a tool introduced by Alfonso et al. (2023). This task preserves the essence of Coller and Williams (1999) with a visual format which has been shown to give better results in terms of consistency in a young population. Students are asked to make 6 decisions, choosing between two options. In option A, the money is delivered in one day, while in option B it is delivered in eight days. Both options start from the same initial payoff, but only option B increases.

For each decision, one variable is assigned starting with *money* and the number of the decision: *money1*, ..., *money6*; (0 = Option A, 1 = Option B). Table 4 shows the percentage of responses for each option, with respect to the subjects who do answer that task.

Table 4: Percentage of responses for each decision

	A	B	n
<i>money1</i>	83.63%	16.37%	5,693
<i>money2</i>	57.80%	42.20%	5,692
<i>money3</i>	46.93%	53.07%	5,689
<i>money4</i>	39.34%	60.66%	5,689
<i>money5</i>	36.66%	63.34%	5,687
<i>money6</i>	29.19%	70.81%	5,686

From these results, the variable *patience* is constructed as the total number of options B chosen by the individual, as well as the normalized one (*patienceN*). Furthermore, a variable called *cns_td* is also included to indicate which individuals show consistent behavior. We consider that an individual shows consistent behavior in time preference elicitation when, once she has chosen option B (receive in eight days instead of tomorrow), she does not choose option A in the following decisions. In other words, if they are consistent, they do not switch back.

[Incluir más detalles, capturas de la tarea?]

3.5 Risk preferences

Similar to eliciting preferences for time discount, we use a visual task to measure risk aversion based on Holt and Laury (2002). According to the results reported by Vasco and De Francisco (2023), this task (*The Gumball Machine*) is best suited in contexts with non-standard subjects. Students have to make 6 decisions, each with two options presented with gumball machines containing different payoffs and different probabilities. Option A is the safer alternative and Option B is the riskier one.

³The option to not answer personal questions and continue forward was available.

For each decision, a variable is associated with the name prob followed by the number of the decision: *prob1*, ..., *prob6*; (0 = Option A, 1 = Option B). Table 5 shows the proportion of responses.

Table 5: Percentage of responses for each decision

	A	B
<i>prob1</i>	90.40%	9.60%
<i>prob2</i>	82.32%	17.68%
<i>prob3</i>	45.30%	54.70%
<i>prob4</i>	17.97%	82.03%
<i>prob5</i>	9.55%	90.45%
<i>prob6</i>	4.67%	95.33%

The level of risk is calculated by simply adding up all the decisions that option B has been chosen, *risky*, this variable has also been normalized (*riskyN*). In order to differentiate individuals who show consistent behavior, we focus on recognizing three types of inconsistencies (Vasco and De Francisco, 2023). The *dominated1* variable refers to individuals who choose option A, the dominated option, in the first decision. While the *dominated6* variable collects information on individuals who choose option B, the dominated option, in the last decision. In addition, we also include a variable relating to any switch back in the set of decisions made by the individual, *swb_risk*. Finally, the variable *cns_risk* identifies those individuals who show consistent behavior.

3.6 Social preferences

For testing inequality aversion in teenagers we used two experimental designs.

The first one was based on Fehr et al. (2008). We built three binary decisions where subjects must decide between an altruistic and an individualistic choices. These decisions were presented in the same order for one part of the sample, and it was randomized for other part of the subjects [PONER UNA NOTA A PIE DE PÁGINA O UNA TABLA QUE DIGA CUÁNTOS SUJETOS TIENEN CADA DISEÑO. Y DENTRO DEL PRIMERO, QUIÉN LO HIZO ALEATORIZADO Y QUIÉN NO].

This first experimental design included:

- Q1** What do you prefer?
A) 10€ for you and 10€ for the other person; B) 10€ for you and 0€ for the other person (*socpref1*)
- Q2** What do you prefer?
A) 10€ for you and 10€ for the other person; B) 10€ for you and 20€ for the other person (*socpref2*)
- Q3** What do you prefer?
A) 10€ for you and 10€ for the other person; B) 20€ for you and 0€ for the other person (*socpref3*)
- [HAY QUE PONER LOS VALORES DE SOCPREF, SI ES 1 QUÉ SIGNIFICA Y SI ES 0 QUÉ SIGNIFICA].

The second experimental design was based on Corgnet et al. (2015) and Brañas-Garza et al. (2022). It was a version extended of the first one, where subjects must decide between two binary choices in six questions. These decisions were always presented in the same order, and it included:

- Q1** What do you prefer?:
A) 1€ for you and 1€ for the other person; B) 0.8€ for you and 1.6€ for the other person (*sp1*)
- Q2** What do you prefer?
A) 1€ for you and 1€ for the other person; B) 1.2€ for you and 0.4€ for the other person (*sp2*)
- Q3** What do you prefer?
A) 1€ for you and 1€ for the other person; B) 1€ for you and 1.8€ for the other person (*sp3*)
- Q4** What do you prefer?
A) 1€ for you and 1€ for the other person; B) 1€ for you and 0.6€ for the other person (*sp4*)

- Q5** What do you prefer?
 A) 1€ for you and 1€ for the other person; B) 1.6€ for you and 0.4€ for the other person (*sp5*)
- Q6** What do you prefer?
 A) 1€ for you and 1€ for the other person; B) 1.1€ for you and 1.9€ for the other person (*sp6*)
- [HAY QUE PONER LOS VALORES DE SOCPREF, SI ES 1 QUÉ SIGNIFICA Y SI ES 0 QUÉ SIGNIFICA].

3.7 Probabilistic beliefs

We test their probability knowledge with an adaptation of Estepa et al. (2021) based on the approach of Delavande and Kohler (2009). The task involves specifying between 0 and 100⁴ the probability of an event actually occurring. We include the following seven questions:

- Q1** Imagine I have a basket with 5 apples: 1 green and 4 red. If I ask you to pick one of the apples without looking at the inside of the basket, how likely (from 0 to 100) do you think you will pick the green apple? (*delavande_apple20*)
- Q2** Imagine I have a basket with 10 apples: 1 green and 9 red. If I ask you to pick one of the apples without looking at the inside of the basket, how likely (from 0 to 100) do you think you will pick the green apple? (*delavande_apple10*)
- Q3** How likely (from 0 to 100) do you think you will eat rice in the next week (including today)? (*delavande_riceweek*)
- Q4** How likely (from 0 to 100) do you think you will eat rice in the next month (including today)? (*delavande_ricemonth*)
- Q5** How likely (from 0 to 100) do you think you are not going to attend school during the entire next month (including today)? (*delavande_sch0*)
- Q6** How likely (from 0 to 100) do you think you are going to take a shower at least once in the next month (including today)? (*delavande_bath100*)
- Q7** How likely (from 0 to 100) it is that you will go to university? (*delavande_uni*)

We also tested these questions with two treatments: numerical values and sliders. In the first one subjects must answer typing the number they thought, while in the second one they must move a slider to the value they wanted to answer.

[PONER UNA TABLA CON EL NÚMERO DE OBSERVACIONES QUE TIENEN SLIDER Y NUMERICAL]

3.8 Abilities

We included two tasks to measure skills: the Cognitive Reflection Test (CRT) and some mathematical-financial questions (FinAb). Subjects always answered both tasks in the same order: first they responded CRT and then FinAb. Both tasks were included in that fixed order, although the questions in each task were displayed randomized.

On the one hand, in the same vein as Thomson and Oppenheimer (2016), we include three adapted CRT questions:

- Q1** Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?
- Intuitive answer: June, Correct answer: Emily
 - Reflexive variable *crt1*: 0 = Other, 1 = Emilia
 - The original response is included in *emily*.
- Q2** In a library, the number of books doubles every month. If the library takes 48 months to fill, how long will it take to fill it halfway?
- Intuitive answer: 24, Correct answer: 47
 - Reflexive variable *crt2*: 0 = Other, 1 = 47
 - The original response is included in *library*.
- Q3** If you are running a race and you pass the person in second place, what place are you in?
- Intuitive answer: 1, Correct answer: 2
 - Reflexive variable *crt3*: 0 = Other, 1 = 2
 - The original response is included in *race*.

On the other hand, FinAb includes basic operations and interest rates⁵:

⁴Values outside this range are denoted by 999

⁵Those who refused to answer or did not know the answer could enter 99

- Q1** If there are 5 people holding a winning lottery ticket and the price to be distributed is 2 million euros, how much money will each person receive?
- Correct answer: 400,000
 - *fin1*: 0 = Incorrect, 1 = Correct
 - The original response is included in *lottery*.
- Q2** Imagine you have 100 euros in a savings account. The account is earning interest at an annual rate of 10%. How much will you have in the account after two years?
- Correct answer: 121
 - *fin2*: 0 = Incorrect, 1 = Correct
 - The original response is included in *bank1*.
- Q3** Imagine you have 100 euros in a savings account and the annual interest rate you earn on your savings is 2%. If you keep the money in the account for 5 years, how much money will you have at the end of 5 years? Multiple choice answer: A) Less than 102 €, B) Exactly 102 €, C) More than 102 €, D) I don't know (99).
- Correct answer: C) More than 102 €
 - *fin3*: 0 = Incorrect, 1 = Correct
 - The original response is included in *bank2*.

3.9 Networks

We elicit four social networks for each year, two positives (friends and best friends) and two negatives (enemies and worst enemies). We displayed a list of students in the same year, including other groups, and we asked participants to tick their friends on the first screen. They are then asked to report who their best friends are from the students they have selected as friends. The next screen showed the same list but asked them to tick their enemies⁶. Finally, they are also asked to select the worst relationships, from those that they have identified as bad relationships.

In addition, we have included new questions related to beliefs. Subjects have to predict who from the list shown will select them as friends, best friends, enemies or worst enemies.

Other types of questions related to the position in the network are included. We asked about the class delegate. They also report who they think is the most popular person and the most connected person.

- Q1** Please select your friends (*friend*)
- Q2** Please select your best friends (*friend2*)
- Q3** Please select those with whom you do not have a good relationship (*enemy*)
- Q4** Please select those with whom you have an especially bad relationship (*enemy2*)
- Q5** Who do you think will select you as a friend? (*sfriend*)
- Q6** Who do you think will select you as a best friend? (*sfriend2*)
- Q7** Who do you think will select you because you don't have a good relationship with that person? (*senemy*)
- Q8** Who do you think will select you because you have an especially bad relationship with that person? (*senemy2*)
- Q9** Who is the delegate of your classroom? (*delegate*)
- Q10** Who do you think will be the person most often identified as a best friend? (*friend3*)
- Q11** If you would like to send a message and you would like it to spread to the whole class, but you could only contact one person, who would you choose? (*message*)

Once we have the information about the interactions reported by the students, we build the social networks and proceed to the social network analysis. We provide both individual and global measures of each network.

- **Individual network measures:**
 - Centrality: Degree, In degree, Out degree, Betweenness, Closeness, Eigenvector
 - Reciprocity: Reciprocal degree
 - Segmentation: Clustering
- **Global network measures:** No. nodes, No. edges, Density, Global reciprocity, Assortativity, Average in degree, standard deviation of in degree, Average degree, Standard deviation of degree, Global clustering, Average clustering, Standard deviation of clustering, Gigant component, No. isolates, Mean distance, Diameter, Gender homophily, Modularity, No. communities, No. important communities.

⁶We never used the term enemy, we rather asked about classmates who aren't friendly or don't have a good feeling for each other.

3.10 Bullying

We include a special section on school bullying. First, we present them a screen with the description. And in the next screen, we ask them to identify the cases they know (max. 3).

Definition of school bullying: *"It happens when, usually, several students intentionally disturb a classmate, who does not manage to stop the behavior. It may involve one or more of the following acts: Teasing, name-calling, bad-mouthing, rejection, negative comments and humiliation, either face-to-face or online (including social media networking (Facebook, WhatsApp, Twitter...), threats, hitting, pushing and similar."*

As for the network formation, we present them the list with all the classmate names, including themselves. Therefore, we capture two different bullying variables with this design, whether they self-identify as a victim (***self_bullying***) or whether others identify them as a victim (***others_bullying***). Both variables are binary, they take a value of 1 if the subject is a victim and 0 otherwise. Note that if the subject has not answered the question, ***self_bullying*** will be a missing value, but if she has been mentioned by a peer, that value will appear in ***others_bullying***. In addition, we include another numerical variable that counts how many times a victim is mentioned as such by her peers (***n_others_bullying***). Lastly, we do not ask for the name of the bully.

Table 6: Bullying cases

	Reported by peers	
	Yes	No
Self-reported	Yes 40	43
	No 312	2,991

Note: There are 682 missing values that we do not have information reported by the subjects themselves.

3.11 Strategical thinking

3.12 Creativity

4 Discussion and closing remarks

References

- Alfonso, A., Brañas-Garza, P., Jorrat, D., Lomas, P., Prissé, B., Vasco, M., & Vázquez-De Francisco, M. J. (2023). The adventure of running experiments with teenagers. *Journal of Behavioral and Experimental Economics*, 102048. <https://doi.org/10.1016/j.socec.2023.102048>
- Anders, J., & Micklewright, J. (2015). Teenagers' expectations of applying to university: How do they change? *Education sciences*, 5(4), 281–305.
- Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of health economics*, 27(5), 1260–1274.
- Belzil, C., & Leonardi, M. (2013). Risk aversion and schooling decisions. *Annals of Economics and Statistics/Annales d'économie et de statistique*, 35–70.
- Bettinger, E., & Slonim, R. (2007). Patience among children. *Journal of Public Economics*, 91(1-2), 343–363.
- Brañas-Garza, P., Cabrales, A., Espinosa, M. P., & Jorrat, D. (2022). The effect of ambiguity in strategic environments: An experiment. *arXiv preprint arXiv:2209.11079*.
- Bryn, N. (2009). Empty archives. *Nature*, 461(7261), 160–163.
- Castillo, M., Ferraro, P. J., Jordan, J. L., & Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, 95(11-12), 1377–1385.
- Chapman, G. B. (2003). Time discounting of health outcomes.
- Chesson, H. W., Leichliter, J. S., Zimet, G. D., Rosenthal, S. L., Bernstein, D. I., & Fife, K. H. (2006). Discount rates and risky sexual behaviors among teenagers and young adults. *Journal of Risk and uncertainty*, 32, 217–230.
- Coller, M., & Williams, M. B. (1999). Eliciting individual discount rates. *Experimental Economics*, 2(2), 107–127.
- Corgnet, B., Espin, A. M., & Hernán-González, R. (2015). The cognitive basis of social behavior: Cognitive reflection overrides antisocial but not always prosocial motives. *Frontiers in behavioral neuroscience*, 9, 287.
- Delavande, A., & Kohler, H.-P. (2009). Subjective expectations in the context of hiv/aids in malawi. *Demographic research*, 20, 817.
- Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3), 1238–1260.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3), 522–550.
- Eckel, C., Johnson, C., & Montmarquette, C. (2005). Saving decisions of the working poor: Short-and long-term horizons. In *Field experiments in economics* (pp. 219–260). Emerald Group Publishing Limited.
- Estepa, L., Jorrat, D., Orozco-Olvera, V., & Rascon Ramirez, E. (2021). *Beans vs slider: Eliciting probabilities in the field* (tech. rep.). Mimeo.
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079–1083.
- Golsteyn, B. H., Grönqvist, H., & Lindahl, L. (2014). Adolescent time preferences predict lifetime outcomes. *The Economic Journal*, 124(580), F739–F761.
- Guiso, L., & Paiella, M. (2004). The role of risk aversion in predicting individual behaviors. *Available at SSRN 608262*.
- Hanushek, E. A., Kinne, L., Lergetporer, P., & Woessmann, L. (2020). *Culture and student achievement: The intertwined roles of patience and risk-taking* (tech. rep.). National Bureau of Economic Research.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3), 411–482.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655. <https://doi.org/10.1257/000282802762024700>

- Kaye, J., Heeney, C., Hawkins, N., De Vries, J., & Boddington, P. (2009). Data sharing in genomics—re-shaping scientific practice. *Nature Reviews Genetics*, 10(5), 331–335.
- Lawless, L., Drichoutis, A. C., & Nayga, R. M. (2013). Time preferences and health behaviour: A review. *Agricultural and Food Economics*, 1, 1–19.
- List, J. A., Petrie, R., & Samek, A. (2023). How experiments with children inform economics. *Journal of Economic Literature*, 61(2), 504–564.
- Mergel, I., Schweik, C. M., & Fountain, J. E. (2009). The transformational effect of web 2.0 technologies on government. *Available at SSRN 1412796*.
- O'Donoghue, T., & Rabin, M. (2001). Risky behavior among youths: Some issues from behavioral economics. In *Risky behavior among youths: An economic analysis* (pp. 29–68). University of Chicago Press.
- Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, 103(1), 510–531.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99–113. <https://doi.org/10.1017/S1930297500007622>
- Vasco, M., & De Francisco, M. J. V. (2023). Holt-laury as a gumball machine. *Available at SSRN 4491386*.