WORKING PAPERS

N° TSE-627

February 2016

"Survey on aspiration and expectations of high school students"

José Raimundo Carvalho and Thierry Magnac

Toulouse School of Economics

SURVEY ON ASPIRATIONS AND EXPECTATIONS OF HIGH SCHOOL STUDENTS

José Raimundo Carvalho

CAEN, Universidade Federal do Ceará

Thierry Magnac

Toulouse School of Economics

First version: November 2014

This version February 19, 2016

Abstract

In this document, we review the main characteristics of the survey undertaken in Ceara in 2014 among students of public and private high schools and regarding their characteristics and behavior relative to the choice of college and undergraduate degrees.

1 Context

Access to Federal universities in Brazil is a very competitive and formal process[1]. In contrast with countries such as the United States, where the predominant selection system uses multiple criteria (grades, letters of recommendation, interview, etc.), selection by examination and objective classification is widespread in Brazil. The recently created centralized matching system SISU (Unified Selection System) came in 2010 as a substitute for the Vestibular, an almost century-old institution that was decentralized at the level of universities. Each Federal university could organize the selection of students through specific exams as they saw fit. Previous work in Carvalho, Magnac, and Xiong (2014) showed that these selection mechanisms were not leading to optimal allocations. This is in this context that the new selection system SISU started and since then has been upheld by the Federal government as a substitute for the Vestibular.

From an economic point of view, the properties of the mechanism governing the allocation of students to specific university undergraduate degrees[2] are of great importance (see for a survey Sönmez and Ünver (2011)). The working of the labor market and the development of a region are not only related to the number of seats opened in each undergraduate degree but also to the quality of the match between students' skills and preferences and undergraduate degrees.

The recent creation of SISU can be understood as the wind-up of the institutional development of Vestibular in Brazil. Over the years the Vestibular experienced many modifications and/or adaptations, especially with regard to the selection criteria and institutional aspects. For instance, the Vestibular at Universidade Federal do Ceara changed criteria as to the format of exams applied to students from a unique homogenous exam in early years, to a unique heterogenous (with major specific exams), and finally ended up with a two-step type of exam where the first step consisted of a unique homogenous exam and the second step had a unique major specific exam. Also,

[2]In Brazilian universities students are not pooled together before they choose different fields or "majors". They have to choose from the beginning, a specific undergraduate degree, even though similar undergraduate degrees share common courses in their starting phases.

the Vestibular started by allowing students to select only one major then moved to a period where students might rank up to 2 or 5 choices of majors, and finally returned to early days with fewer proposed major choices. It is worth mentioning that the Vestibular and universities always required those choices to be made *ex ante*, i.e. before exams are taken and that in Brazil once a student starts a major the likelihood of switching to another major is very limited, demanding almost always a re-taking of an entrance exam.

From an economic perspective, such changes represented attempts to tune in the best allocation mechanism between students and degrees. The implementation of SISU, can be rationalized as an attempt by the Ministry of Education to centralize the matching of students with Federal universities, this time on a national scale. Centralization minimizes the costs of selection of students since it avoids that all universities evaluates every student they are interested in. Nonetheless, it comes at the cost of having a common set of exams of students across universities and thus give more control to the central authorities (see Che and Koh (2014) and Hafalir, Hakimov, Kübler, and Kurino (2014)).

Since students' choice of an undergraduate degree is the result of a rational calculation of expected costs and benefits reflecting individual preferences and opportunity costs, budget constraints and earnings expectations, SISU is very likely to bring about important changes in this process (see Gontijo (2008)). Despite that, few studies up to now (Abreu and Carvalho (2014)) considered thoroughly the question of what affects students' choice of major or the impact of the implementation of SISU although there are connected analyses (Aygün and Bó (2014)). The survey we collected is aimed at understanding these issues about the choice of undergraduate degrees by high school students as well as the allocating role played by SISU.

1.1 The PAEST survey

Collecting detailed information about the context students face during their last high school year is key to analyzing these issues. This is why we developed the socioeconomic survey PAEST (Pesquisa de Aspirações e Expectativas dos Estudantes Concludentes do Ensino Médio - Survey on Aspirations and Expectations of High School Students) and collected information on family structure, socio-demographics, educational achievement and performance, networks through family and friends, sources and quality of information about undergraduate studies and expectations

about labor market and educational choices. The survey received the approval of the Comitê de Ética ($N°$ 12523713.2.0000.5054) regarding privacy issues of information.

The questionnaire (see Appendix A) contains the following sections: Socioeconomic conditions; Performance and educational ability; Networks: Family, friends and reference groups; Sources of information; Educational choices; Preferences over degrees and expectations about undergraduate degrees and labor markets; Other information about risk preferences and recent events. This survey instrument was administered to a sample of students in the third year of high schools in public and private schools of the Northeastern state of Ceará and in particular its capital, Fortaleza, its metropolitan area and other large municipalities within the state.

We hope that PAEST would generate a deeper understanding of how students perceive and act in the process of choosing their undergraduate degree in university. This knowledge can indeed serve to further improvements in the system of allocation, SISU, as the main mechanism of university access in Brazil.

2   Sample Design and Field Work

2.1   Sample Size

The sampling plan was drawn up by stratifying it first in three geographical groups: Fortaleza, the metropolitan area of Fortaleza, denoted RMF in the following, and a group composed of three large municipalities (Sobral, Juazeiro do Norte e Crato) in order to ensure that the sample is statistically representative of marked geographical and socioeconomic differences. The sample size (n) was determined by using the following equation:

$$n = \left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2 \times p \times q \tag{1}$$

Where: $1 - \alpha$ is a level of confidence, $Z_{\frac{\alpha}{2}}$ the $\frac{\alpha}{2}$ quantile of a standard normal random variable, p is a population parameter describing a proportion, $q = 1 - p$, E, an error to be reached for the population parameter, p and n is the resulting sample size. The worst case is obtained when $p = 0.5$.

Table (1) shows the proposed sample sizes for each geographic area in Ceará.

Table 1: Proposed Sample Size

Area	Sample Error	Sample Size
Fortaleza	3%	1,000
RMF	4%	600
Interior	4%	600
TOTAL		2,200

There are huge differences between public (free of charge) and private high schools in Brazil, mainly regarding the quality of installations, teachers and teaching. Overall, private schools are of better quality compared to their public counterparts and this leads to selection issues. This is why we stratified further our sample between private and public schools as to reflect as closely as possible the composition of the population. Table (2) below shows the proposed sample stratification between public and private schools:

Table 2: Public *versus* Private Stratification

Area	Total	Particular	Public
Fortaleza	1,000	250	750
RMF	600	50	550
Interior	600	100	500
TOTAL	2,200	400	1,800

Last but not least, our actual sample unit is a whole school class (and not individual students) since one of our interests is to investigate the rôle of peers and networks that might impact the choice of degrees by students.

Table (3) shows the final stratification proposed. The first three columns refer to number of students while the last three columns refer to classes assuming that on average classes have 40 students (see DATAINFO (2015)).

Table 3: Final Stratification

AREA	MUNICIPALITY	Total	Private	Public	Total	Private	Public
Fortaleza	Fortaleza	1,000	250	750	33	8	24
RMF	Caucaia	300	50	250	10	2	8
RMF	Maracanaú	240	0	240	8	0	8
RMF	Eusébio	60	0	60	2	0	2
Interior	Sobral	240	80	160	9	4	6
Interior	Juazeiro do Norte	240	20	220	9	1	8
Interior	Crato	120	0	120	4	0	4
TOTAL		2,200	400	1,800	75	15	60

2.2 Selection Procedure for Schools and School Classes

For public schools, we used a list of public schools from an administrative database provided by the State Department of Education - SEDUC. Schools are then stratified by municipality and are randomly sampled. Classes to be interviewed in each school were chosen randomly whenever possible except when some external factor prevented it. For instance, some public schools had less than two available classes; or a previously chosen school announced close to the day of interviews that some classes were not available due to absence of teachers of unexpected outside activities; or finally, the class had a much lower than expected number of students than previously disclosed. Overall, those issues were minor.

Altogether 36 public schools were visited. The apparent discrepancies between this number and that appearing at the bottom of the last column of Table 3 (i.e. 60) needs some clarification. We realized after sampling design, but before actual sampling, that morning and afternoon classes do have some differences that might be important to the survey. Hence, to include this new stratification and keep logistics manageable we decided to sample at least one morning and one afternoon class at each school, whenever possible. So, the 60 appearing in Table 3, last column and row, should be read as the total number of classes to be sampled in public school. We visited 36 public schools, sampled two classes (a morning and a afternoon) for the vast majority of them, and for very few schools we sampled either one or more than two classes.

6

For private schools, the choice was made for convenience of the schools, although we are able to include schools which exhibit variation in size and/or quality. Contacts were made by telephone and face to face with various schools, and after presenting the project to the boards and getting their approval we ended with 11 private schools participating. Again, discrepancies between this number and that appearing at the bottom of the seventh column of Table 3 (say, 15) should be read similarly to that of public schools.

2.3   Implemented Sample Size

The final sample changed slightly in size with respect to what was anticipated. The final database includes 2,342 interviewed students, 142 in excess of what was originally planned. The lower bounds for students in both private and public schools were achieved (see Table 3, fourth and fifth columns of the bottom line).

Table 4: Public *versus* Private Stratification - Final

Area	Total	Particular	Public
Fortaleza	1,051	244	807
RMF	604	79	525
Interior	687	189	498
TOTAL	2,342	512	1,830

2.4   Field Work

Field work was done in two steps: a pre-test was administered in November and December 2013 and the final questionnaire was administered between 4 August to 19 September 2014. The schedule of visits to schools was:

1. August 4th to September 29th, 2014 - Fortaleza

2. 2nd to 18th of September, 2014 - Metropolitan region

3. August 18 to September 22, 2014 - Other municipalities

More specifically, the following procedures were followed:

1. In each public school, we randomly sampled two classrooms, one morning and the other afternoon. The traditional distinction between morning and afternoon classes applies all over the country. For private schools the protocol was similar, except that there is no randomization.

2. About a week prior to administering the questionnaire, a representative of our research group presented details about the project to students. Each student had to give written consent to participating in the survey;

3. At a previously scheduled date, a team of interviewers carried out the survey at each participating school.

4. The entire procedure was accompanied by a representative of the research group.

3 A Brief Overview of Results

3.1 Basic Demographics

Results are as follows[3]. As shown by Table 5 58.3% of students are girls. Age distribution appears in Table 6 where the mode of 17 years is the norm for Brazilian standards.

Table 5: Sex

	Frequency	Percent	Valid Percent	Cumulative Percent
Male	973	41.5	41.6	41.6
Female	1367	58.3	58.4	100.0
99999	1	.0	.0	100.0
Total	2340	99.8	100.0	
System	4	.2		
	2344	100.0		

Table 7 presents a typical Brazilian mixed race country picture: 63.5% of the students report themselves as brown (**"pardo"**) while 25.7% considered themselves as white and 7.5% as black.

_____

[3]In all coming Tables, a value of 77777 means a **Do not know/Did not answer**; 88888 means a **Non-applicable answer** and 99999 is a true **missing observation**. Also, the missing reference for IBM SPSS, say _System_, refers to two observations added after data compilation and, for all purposes, can be interpreted as a 99999.

Table 6: Age

	Frequency	Percent	Valid Percent	Cumulative Percent
15	6	.3	.3	.3
16	279	11.9	11.9	12.2
17	1288	54.9	55.0	67.2
18	532	22.7	22.7	89.9
19	156	6.7	6.7	96.5
20	45	1.9	1.9	98.5
21	15	.6	.6	99.1
22	9	.4	.4	99.5
23	1	.0	.0	99.5
25	2	.1	.1	99.6
26	3	.1	.1	99.7
27	1	.0	.0	99.8
28	1	.0	.0	99.8
36	2	.1	.1	99.9
43	1	.0	.0	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 7: Race

	Frequency	Percent	Valid Percent	Cumulative Percent
White	602	25.7	25.7	25.7
Black	175	7.5	7.5	33.2
Pardo	1489	63.5	63.6	96.8
Asian	53	2.3	2.3	99.0
Brazilian native	20	.9	.9	99.9
77777	2	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.2 Family Structure

As regards family structure, 58.4% of students report that they are living with their parents while 22.8% are living with their mother only (see Table 8).

Table 8: Co-Habitation

	Frequency	Percent	Valid Percent	Cumulative Percent
With your father and your mother	1370	58.4	58.5	58.5
With your father only	68	2.9	2.9	61.4
With your mother only	534	22.8	22.8	84.2
With other relatives	343	14.6	14.6	98.8
Alone	10	.4	.4	99.3
With a roommate	12	.5	.5	99.8
77777	4	.2	.2	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

In Table 9 36.0% of students report that the total monthly income of their parents (or legal guardians) varies from one to two minimum wages while for 16.8% of them household income varies between two and three minimum wages. Note that income is the sum of all possible sources of income such as earnings, pension, retirement, social benefits, rents and so on.

Table 9: Parents Total Income

	Frequency	Percent	Valid Percent	Cumulative Percent
$\leq$ R\$ 724.00	564	24.1	24.1	24.1
(R\$ 724.00 a R\$ 1,448.00]	843	36.0	36.0	60.1
(R\$ 1,448.00 a R\$ 2,172.00]	394	16.8	16.8	76.9
(R\$ 2,172.00 a R\$ 3,620.00]	211	9.0	9.0	85.9
(R\$ 3,620.00 a R\$ 7,240.00]	147	6.3	6.3	92.2
(R\$ 7,240.00 a R\$ 10,860.00]	44	1.9	1.9	94.1
(R\$ 10,860.00 a R\$ 14,480.00]	24	1.0	1.0	95.1
$\geq$ R\$ 14,480.00	24	1.0	1.0	96.1
77777	88	3.8	3.8	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

In terms of parental education, 27.5% of fathers or stepfathers have completed secondary school while 56.6% of them did not complete primary school (see, Table 10). As to mothers

or stepmothers, 28.2% completed secondary school and 45.8% did not finish elementary school (see, 11). It is interesting to note that among parents with an undergraduate degree, 7.0% of fathers/stepfathers only have a college degree while this percentage for mothers/stepmothers is 10.0%.

Table 10: Father Schooling Level

	Frequency	Percent	Valid Percent	Cumulative Percent
No education	82	3.5	3.5	3.5
Incomplete fundamental school	744	31.7	31.8	35.3
Complete fundametal school	269	11.5	11.5	46.8
Incomplete high school	161	6.9	6.9	53.6
Complete high school	645	27.5	27.5	81.2
Incomplete University/College	73	3.1	3.1	84.3
Complete University/College	179	7.6	7.6	91.9
77777	172	7.3	7.3	99.3
99999	17	.7	.7	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 11: Mother Schooling Level

	Frequency	Percent	Valid Percent	Cumulative Percent
No education	57	2.4	2.4	2.4
Incomplete fundamental school	717	30.6	30.6	33.0
Complete fundametal school	299	12.8	12.8	45.8
Incomplete high school	226	9.6	9.6	55.5
Complete high school	660	28.2	28.2	83.6
Incomplete University/College	84	3.6	3.6	87.2
Complete University/College	235	10.0	10.0	97.3
77777	64	2.7	2.7	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.3 Schooling Decisions

Tables 12 and 13 show, respectively, the type of school students attended in primary and secondary school. In terms of students' educational attainment, 56.9% studied in a public institution throughout the whole elementary (primary) school period. As to the secondary school (2 first years), 74.5% studied in a public school.

Table 12: Primary Schooling: Public *versus* Private

	Frequency	Percent	Valid Percent	Cumulative Percent
All inpublic schools	1334	56.9	57.0	57.0
Mostly in public schools	215	9.2	9.2	66.1
Mostly in private schools	183	7.8	7.8	74.0
All in private schools	608	25.9	26.0	99.9
77777	2	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 13: Secondary Schooling: Public *versus* Private

	Frequency	Percent	Valid Percent	Cumulative Percent
All inpublic schools	1746	74.5	74.6	74.6
Mostly in public schools	54	2.3	2.3	76.9
Mostly in private schools	51	2.2	2.2	79.0
All in private schools	491	20.9	21.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.4 Student's Own Performance Assessment

Students were also requested to assess their performance during high school. Table 14 shows that 46.2% of students considered themselves as being of average quality, 26.5% as above average while 21.6% considered themselves very good during primary school. As to performance during high school (the first 2 years), Table 15 shows 51.2% of students considered themselves as being of average quality, 22.2% as above average while 21.3% considered themselves very good.

Table 14: Performance during Primary School

	Frequency	Percent	Valid Percent	Cumulative Percent
Very good	758	32.3	32.4	32.4
Above average	578	24.7	24.7	57.0
Average	950	40.5	40.6	97.6
Below average	51	2.2	2.2	99.8
Far below average	4	.2	.2	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 15: Performance during Secondary School

	Frequency	Percent	Valid Percent	Cumulative Percent
Very good	499	21.3	21.3	21.3
Above average	520	22.2	22.2	43.5
Average	1201	51.2	51.3	94.8
Below average	112	4.8	4.8	99.6
Far below average	8	.3	.3	99.9
77777	1	.0	.0	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.5   Sources of Information

We asked students questions about their main source of information about academic and professional matters in the last 6 months. Table 16 up to Table 29 provide fourteen possible sources of information that a high school student might refer to while choosing undergraduate majors:

1. Parents (father, mother, stepfather, stepmother, legal guardian)

2. Brothers, sisters

3. Cousins, uncles, aunts, other relatives

4. School friends

5. Other friends

6. Teachers from your school

7. Professionals (educational/professional advisor, counselor, others)

8. Internet (except Social Medias)

9. Books and magazines

10. Visits to Universities

11. University Professors

12. University students

13. Career expo

14. Social medias

It is important to mention that PAEST also asked about the quality of information of each of these fourteen possible sources as well. Indeed, question 30 is phrased like: *How do you evaluate the quality of information obtained when you talked or sought help/information on issues related to what you would do with your academic and professional life after High School? That is, in your opinion, talking to those people and/or obtaining information from those sources let you in what situation regarding what you would do after High School?*. We believe those two set of questions dealing, respectively, with intensity and quality of information can be explored in more detailed models of schooling choice.

Table 16: Source of Information: Parents

	Frequency	Percent	Valid Percent	Cumulative Percent
None	278	11.9	11.9	11.9
Once	284	12.1	12.1	24.0
Twice	349	14.9	14.9	38.9
Three times or more	1427	60.9	60.9	99.8
77777	1	.0	.0	99.9
88888	2	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 17: Source of Information: Brothers/Sisters

	Frequency	Percent	Valid Percent	Cumulative Percent
None	882	37.6	37.7	37.7
Once	294	12.5	12.6	50.2
Twice	325	13.9	13.9	64.1
Three times or more	714	30.5	30.5	94.6
77777	21	.9	.9	95.5
88888	105	4.5	4.5	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

14

Table 18: Source of Information: Cousins, Uncles, Aunts, other Relatives

	Frequency	Percent	Valid Percent	Cumulative Percent
None	748	31.9	31.9	31.9
Once	386	16.5	16.5	48.4
Twice	443	18.9	18.9	67.3
Three times or more	758	32.3	32.4	99.7
77777	3	.1	.1	99.8
88888	3	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 19: Source of Information: School Friends

	Frequency	Percent	Valid Percent	Cumulative Percent
None	132	5.6	5.6	5.6
Once	181	7.7	7.7	13.4
Twice	346	14.8	14.8	28.1
Three times or more	1679	71.6	71.7	99.8
77777	3	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 20: Source of Information: Other Friends

	Frequency	Percent	Valid Percent	Cumulative Percent
None	348	14.8	14.9	14.9
Once	291	12.4	12.4	27.3
Twice	430	18.3	18.4	45.6
Three times or more	1267	54.1	54.1	99.7
77777	3	.1	.1	99.9
88888	2	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 21: Source of Information: Teachers from School

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	619	26.4	26.4	26.4
Once	403	17.2	17.2	43.6
Twice	437	18.6	18.7	62.3
Three times or more	876	37.4	37.4	99.7
77777	4	.2	.2	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 22: Source of Information: Professionals (advisor, counselor, others)

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1055	45.0	45.0	45.0
Once	384	16.4	16.4	61.4
Twice	337	14.4	14.4	75.8
Three times or more	558	23.8	23.8	99.7
77777	5	.2	.2	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 23: Source of Information: Internet (except Social Medias)

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	601	25.6	25.7	25.7
Once	260	11.1	11.1	36.8
Twice	378	16.1	16.1	52.9
Three times or more	1099	46.9	46.9	99.8
77777	3	.1	.1	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 24: Source of Information: Books and Magazines

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	840	35.8	35.9	35.9
Once	377	16.1	16.1	52.0
Twice	369	15.7	15.8	67.7
Three times or more	750	32.0	32.0	99.7
77777	3	.1	.1	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 25: Source of Information: Visits to Universities

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1383	59.0	59.1	59.1
Once	480	20.5	20.5	79.5
Twice	257	11.0	11.0	90.5
Three times or more	218	9.3	9.3	99.8
77777	2	.1	.1	99.9
99999	2	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 26: Source of Information: University Professors

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1654	70.6	70.6	70.6
Once	269	11.5	11.5	82.1
Twice	223	9.5	9.5	91.6
Three times or more	191	8.1	8.2	99.8
77777	2	.1	.1	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 27: Source of Information: University Students

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1026	43.8	43.8	43.8
Once	343	14.6	14.6	58.5
Twice	383	16.3	16.4	74.8
Three times or more	587	25.0	25.1	99.9
77777	1	.0	.0	99.9
99999	2	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 28: Source of Information: Career Expo

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1307	55.8	55.8	55.8
Once	562	24.0	24.0	79.8
Twice	271	11.6	11.6	91.4
Three times or more	197	8.4	8.4	99.8
77777	2	.1	.1	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 29: Source of Information: Social Medias

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	764	32.6	32.6	32.6
Once	286	12.2	12.2	44.8
Twice	314	13.4	13.4	58.2
Three times or more	971	41.4	41.5	99.7
77777	4	.2	.2	99.9
99999	3	.1	.1	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.6 Actual and Future Job Market Activities

Finally, when asked whether working or seeking employment, 25.3% of students reported that they were working , while 19.0% did not work, but were looking for a job. However, when asked what they want do after school graduation, 75.6% answered that they had the intention to both work and study, 17.2% just to keep studying and only 3.8% wanted to work only.

Table 30: Labor Market Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes, I work	445	19.0	19.0	19.0
I do not work but I am looking for a job	593	25.3	25.3	44.3
I am not looking for a job	1298	55.4	55.4	99.7
77777	5	.2	.2	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 31: Activity after High School Graduation

	Frequency	Percent	Valid Percent	Cumulative Percent
Studying only	402	17.2	17.2	17.2
Working and studying	1773	75.6	75.7	92.9
Working only	89	3.8	3.8	96.7
Other activities than working or studying	17	.7	.7	97.4
77777	61	2.6	2.6	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.7 Subjective Expectations, Counterfactuals and Risky Behavior

3.7.1 Subjective Expectations and Counterfactuals

Economists recently devoted substantial efforts in order to understand the key role expectations have on economic choices. In order to achieve that understanding subjective data on expectations have to be collected, modeled and interpreted as diligently as when handling revealed preference or "objective" economic data. Subjective data can help disentangling the interplay of preferences, expectations, and opportunities present in the process of choice, and solving identification problems (see the seminal paper of Manski (2004)).

PAEST also asked questions that enable researchers to use subjective expectations about counterfactuals. For instances, question 48 asks *Suppose you would finished High School and would not study anymore. What is the probability (0% to 100%) that you would be employed at age 30?*, and question 52 asks *Suppose you would complete your first option of undergraduate degree and would not study anymore. What is the probability (0% to 100%) that you would be employed at age 30?*. Tables 32 and 33 present those two counterfactuals, respectively.

Table 32: Probability to be Employed **graduating from** High School Only

	Frequency	Percent	Valid Percent	Cumulative Percent
0	16	.7	.7	.7
1	1	.0	.0	.7
2	1	.0	.0	.8
3	2	.1	.1	.9
4	3	.1	.1	1.0
5	13	.6	.6	1.5
10	63	2.7	2.7	4.2
12	2	.1	.1	4.3
15	14	.6	.6	4.9
20	94	4.0	4.0	8.9
25	27	1.2	1.2	10.1
30	181	7.7	7.7	17.8
35	9	.4	.4	18.2
40	184	7.8	7.9	26.0
43	1	.0	.0	26.1
45	12	.5	.5	26.6
47	1	.0	.0	26.6
48	2	.1	.1	26.7
50	492	21.0	21.0	47.7
52	1	.0	.0	47.8
55	6	.3	.3	48.0
58	1	.0	.0	48.1
60	172	7.3	7.3	55.4
65	6	.3	.3	55.7
66	1	.0	.0	55.7
68	1	.0	.0	55.8
70	184	7.8	7.9	63.6
75	19	.8	.8	64.4
77	1	.0	.0	64.5
80	240	10.2	10.2	74.7
83	1	.0	.0	74.8
85	15	.6	.6	75.4
90	176	7.5	7.5	82.9
95	16	.7	.7	83.6
97	1	.0	.0	83.6
98	3	.1	.1	83.8
99	7	.3	.3	84.1
100	341	14.5	14.6	98.6
99999	32	1.4	1.4	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

Table 33: Probability to be Employed **graduating after** Undergraduate First Option

	Frequency	Percent	Valid Percent	Cumulative Percent
3	1	.0	.0	.0
9	1	.0	.0	.1
10	3	.1	.1	.2
20	2	.1	.1	.3
25	3	.1	.1	.4
30	4	.2	.2	.6
35	1	.0	.0	.6
40	12	.5	.5	1.2
45	1	.0	.0	1.2
50	80	3.4	3.4	4.6
55	2	.1	.1	4.7
60	56	2.4	2.4	7.1
62	1	.0	.0	7.1
65	3	.1	.1	7.3
70	128	5.5	5.5	12.7
75	15	.6	.6	13.4
79	1	.0	.0	13.4
80	314	13.4	13.4	26.8
85	25	1.1	1.1	27.9
87	2	.1	.1	28.0
88	1	.0	.0	28.0
89	1	.0	.0	28.1
90	360	15.4	15.4	43.4
91	1	.0	.0	43.5
92	2	.1	.1	43.6
95	53	2.3	2.3	45.8
97	2	.1	.1	45.9
98	16	.7	.7	46.6
99	27	1.2	1.2	47.7
100	920	39.2	39.3	87.0
88888	286	12.2	12.2	99.2
99999	18	.8	.8	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

3.7.2 Risky Behavior

To improve our understanding about how people behave in situations that may be considered risky, we collected data on risky behavior that might be familiar to teenagers. Students are asked about their own willingness to take risks. Such self-assessment complements traditional lottery-type elicitation of risk attitudes. For instance, a quite familiar issue about teenager sexual behavior concerns unprotected sexual relations. Accordingly to Table 34, 72.3 % of sampled individuals reported a probability of unprotected sex close to zero. At the other extreme, 3.0 % claims to be certain of having unprotected sex.

Table 34: Probability of Risky Behavior: Unprotected Sexual Relation

	Frequency	Percent	Valid Percent	Cumulative Percent
Not likely to occur	1703	72.7	72.7	72.7
Very little likelihood to occur	254	10.8	10.8	83.6
Unlikely to occur	172	7.3	7.3	90.9
Moderate probability to occur	83	3.5	3.5	94.4
Very likely to occur	51	2.2	2.2	96.6
Very much likely to occur	30	1.3	1.3	97.9
Certainly to occur	48	2.0	2.0	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

As regards practicing risky sports like parachutes, bungee jumping, and others, there is a considerable percentage of those youths (10.2%) claiming that they would do that for sure! There are eight additional questions in PAEST that survey students behavior concerning risky activities ranging from "Lending to a relative a monetary amount equal to R\$ 300", i.e. around €100; "Wear skimpy or unconventional clothes" up to "Admitting that your preferences are different from your friends preferences".

Table 35: Probability of Risky Behavior: Practicing any Risky Sport

	Frequency	Percent	Valid Percent	Cumulative Percent
Not likely to occur	929	39.6	39.7	39.7
Very little likelihood to occur	277	11.8	11.8	51.5
Unlikely to occur	244	10.4	10.4	61.9
Moderate probability to occur	255	10.9	10.9	72.8
Very likely to occur	246	10.5	10.5	83.3
Very much likely to occur	152	6.5	6.5	89.8
Certainly to occur	238	10.2	10.2	100.0
99999	1	.0	.0	100.0
Total	2342	99.9	100.0	
System	2	.1		
	2344	100.0		

4 Web Survey about Choice of Majors in the SISU

4.1 Motivation

One of the main objectives of our research agenda is to model students' choice of major under the recently created centralized matching system SISU (Unified Selection System) in Brazil. As we mentioned before, collecting detailed information about the context students face during their last high school year is key to analyzing these issues and that justifies the development of PAEST. However, we need to track students choices when they reveal their preferences about majors in SISU. Public institutions possessing administrative data about students' behavior in SISU have been reluctant at disclosing data. To bypass that, we came up with the idea of building a virtual SISU's system and collect students' simulated choices of majors by means of a follow up web survey applied to those who participated in the original PAEST sample.

4.2 A Glimpse at SISU

The SISU system (Unified Selection System) is a complex centralized mechanism designed to allocate university available seats to students. It started to operate in Brazilian universities around 2010. Accordingly to Abreu and Carvalho (2014), SISU selection mechanism is divided into four stages, say, Registration, First Call, Second Call and Waiting List. The following describes each of these steps:

1. **Registration** Only participants in the last ENEM (high school national exam in Brazil) can

register in SISU. The registration period comprises an interval of four days (from Monday through Thursday) during which students can at any time access the system platform (usually located at `www.sisu.mec.gov.br`) and choose, in order of preference, up to two major options offered by institutions participating in the process[4]. At midnight of Monday, SISU makes use of each student's major options, computes the match and at 2:00 am of Tuesday grants students access to the computed match. At this moment, students have detailed information on each of their own rank among options submitted as well as their distance from each option minimum threshold score (based on ENEM scores). This sequence of allowing students to submit their options, computing matches and granting access to information ends at midnight of Thursday, where the last student's major choices become definitive.

2. **First Call** At the end of the registration period, i.e. Thursday midnight, SISU computes a match based on student's definitive major options and make offers to students. Abreu and Carvalho (2014) argue that the mechanism used by SISU to generate these proposed matches to students, at this stage, is the deferred acceptance with students proposing mechanism (see Sönmez and Ünver (2011)). Each student who receives an offer, decides to either **accept** or **reject** his/her offer. Every student who, at this stage, does not receive any offer, wait for potential offers at later stages.

3. **Second Call** For all majors that finished the **First Call** with vacancies, SISU performs a re-offer of those available places. For each re-offered major (call it *major X*), only students with the highest scores among those who had chosen *major X* before can be potentially offered a seat. Besides that, a student must have not received a previous offer of either *major X* nor of a major that was ranked better than this re-offered *major X*. Again, each student who, at this stage, receives an offer, decide to either **accept** or **reject** his/her offer. Every student who, at this stage, does not receive any offer, wait for potential offers at the last stage, i.e., the Waiting List.

4. **Waiting List** All those whose application has not been contemplated **using** his/her first ranked major option so far, are put by SISU on a waiting list. Those who accept to be

---

[4]During the registration period, students can change as many times as desired their options being proposed, since only the existing choices by Thursday midnight will be considered as student's final preference submission.
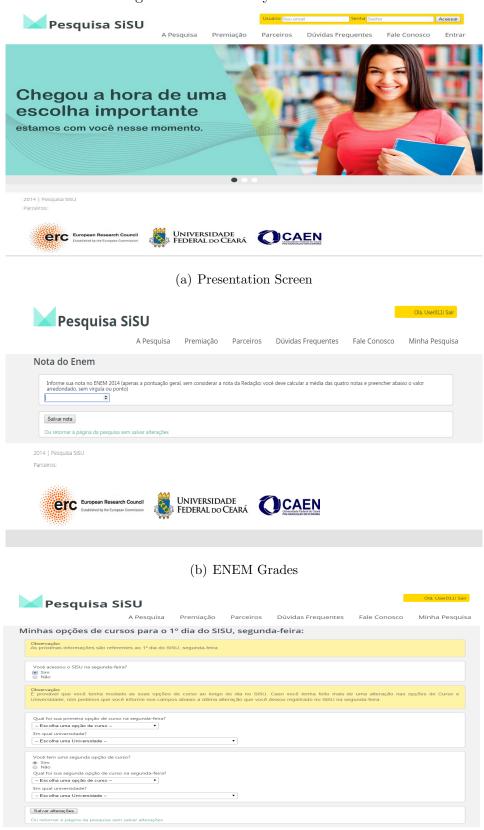
on this waiting list compete exclusively with others having chosen their first ranked major option. That waiting list ends the SISU mechanism.

4.3 Methodology of the Web Survey

Due to the complexities of the SISU and to the fact that the **Registration** phase reveals a lot about how students choose majors, we decided to ask students about their grades at ENEM and to gather information about behavior at the SISU Registration phase. We hired the same firm that conducted the PAEST, DATAINFO (`http://www.datainfo.srv.br/`), and they offered a solution to collect data by means of a web survey.

DATAINFO developed a web based application to collect data in a daily basis about students major choices. In order to improve the quality of the sampled information, the web application was available for access from 19th to 27th, January 2015. This period nests the actual week when SISU took place in 2015, i.e. 19th to 22nd, January 2015. We expected students to be able to reproduce their final major choices in a daily basis in our web application. The survey firm made a campaign advertising about the web application during the month before the actual implementation of the web survey. On the eve of the week of the actual SISU, DATAINFO sent emails to survey participants, reminding them about the web survey and the fact that they were entitled to win prizes upon participation. The email informed how they could access the online environment to fill in the necessary information. The online environment was developed to be as friendly and as easy to follow by students as possible. Figures 1 and 2 present 6 print screens from the actual online environment.

Figure 1: Web Survey Print Screens

(a) Presentation Screen

(b) ENEM Grades

(c) Majors Choice

Figure 2: Web Survey Print Screens

(a) Threshold Grades

(b) End of Survey

(c) Data Confirmation

After the end of SISU's online survey, participants who did not respond to the web environment for any reason had telephone interviews. Such procedure is quite common in web surveys, since this type of survey has typically low response rates (see Dillman, Smyth, and Christian (2014)). Table 36 shows that we succeeded in collecting 820 (394 + 426) cases with valid answers about majors' choice and college options. Students who answered through the online environment have complete data for all four days of SISU; while those students who answered by telephone have data on the last two days of SISU only.

Table 36: Web and Phone Survey Participation

Category		Total	%
Web Survey	Valid questionnaire	394	16.81
	Did not participate in SISU	226	9.64
Phone Survey (valid interviews)	Did not take ENEM	77	3.28
	Valid answer	426	18.17
	Switched off phone or person did not answer	686	29.27
	Wrong number	180	7.68
Phone Survey(invalid interviews)	No phone	48	2.05
	Others	307	13.10
		2344	**100.00%**

5   Accessibility

Data availability is an important issue in the context of our research and we would like to make the data accessible to all researchers. Hence, beginning in August, 1st 2016 we will make available upon written request by email to any student or researcher the full database, under a signed commitment that the data set must be used for scientific purposes only and that there would be no data transfer to other persons without them undergoing the same procedure. In other to abide to privacy rules we will create unique identifiers for each student, each school and each class, so as to preserve the identities of people and schools. For further information, please, contact José Raimundo Carvalho (josecarv@ufc.br) or Thierry Magnac (thierry.magnac@tse-fr.eu).

A Questionnaires

A.1 Questionnaire in Portuguese

The original questionnaire can be downloaded at `http://www.caen.ufc.br/attachments/article/219/paestPORT.pdf`

A.2 Questionnaire in English

This questionnaire can be downloaded at `http://www.caen.ufc.br/attachments/article/219/paestENG.pdf`

B Dictionary of the data

The Dictionary (Codebook in IBM SPSS) can be downloaded at `http://www.caen.ufc.br/attachments/article/219/Dicionario_PAEST.pdf`

References

ABREU, L. C. M., AND J. R. CARVALHO (2014): "Mecanismos de seleção de Gale-Shapley dinâmicos em Universidades Brasilieiras; $SISU$, $SISU_\alpha$ e $SISU_\beta$," Estudos Econômicos CAEN 8, CAEN/UFC, Ceara, Brazil.

AYGÜN, O., AND I. BÓ (2014): "College Admission with Multidimensional Privileges: The Brazilian Affirmative Action Case," Discussion paper, mimeo.

CARVALHO, J. R., T. MAGNAC, AND Q. XIONG (2014): "College choice allocation mechanisms: structural estimates and counterfactuals," Discussion Paper 8550, IZA, Bonn, Germany.

CHE, Y.-K., AND Y. KOH (2014): "Decentralized College Admissions," *Available at SSRN 2399743*.

DATAINFO (2015): "Coleta e Análise Estatística Primária sobre os Determinantes Socioe-
conômicos das Escolhas do Curso Universitário e Comportamento no Sistema de Seleção Unifi-
cada (SISU) dos Estudantes de Ensino Médio do Estado do Ceará," Final Report.

DILLMAN, D. A., J. D. SMYTH, AND L. M. CHRISTIAN (2014): *Internet, phone, mail, and
mixed-mode surveys: the tailored design method.* John Wiley & Sons.

GONTIJO, M. F. (2008): "Uma aplicação da teoria dos jogos ao mercado do vestibular brasileiro,"
Master's thesis, Universidade de São Paulo.

HAFALIR, I. E., R. HAKIMOV, D. KÜBLER, AND M. KURINO (2014): "College admissions with
entrance exams: Centralized versus decentralized," Discussion paper, WZB Discussion Paper.

MANSKI, C. F. (2004): "Measuring expectations," *Econometrica*, pp. 1329–1376.

SÖNMEZ, T., AND M. U. ÜNVER (2011): "Matching, allocation, and exchange of discrete re-
sources," *Handbook of Social Economics*, 1, 781–852.