Pre analysis plan: An experiment on adolescents fairness consideration for admission to education

Siv-Elisabeth Skjelbred

1 Introduction

The criteria used to determine who is qualified for higher education and how students are selected among qualified applicants have large and lasting consequences. There are substantial impacts later in life of how much education you take (e.g., Bhuller et al., 2017), where you study (e.g., Cohodes and Goodman, 2014; Bleemer, 2022; Canaan and Mouganie, 2018) and what you study (e.g., Hastings et al., 2013; Kirkebøen et al., 2016; Bleemer and Mehta, 2022). A recurrent political debate in many countries is the fairness of the criteria used to select between applicants. To design a fair admission system one needs to determine what fairness in education is, or should be. An underlying conflict in the debate about the fairness of admission to education is that different fairness ideals professes different views on which factors individuals should be held accountable for.

This pre-analysis plan presents the data sources, the structure of an experiment, and the empirical strategy of a study focusing on this issue.

2 Research strategy

The experiment is conducted as a part of a survey to 10th graders in the municipality of Oslo, Norway. The topic of the survey is application to upper secondary education. All schools in the municipality of Oslo has been invited to participate, and 9 schools had agreed at the deadline. A reminder was sent out to the schools after the pre-analysis plan was submitted, hence the final number of schools might be higher. All 10th grade students in ordinary classes in these schools will be invited to take the survey during school hours.

The research project will be implemented in February 2024. The pre-analysis plan was completed before the research project was implemented and the researchers

did not have access to the data set before the plan was registered at the AEA RCT trial.

3 Design

We plan to conduct a between subject treatment manipulation. We are investigating how and whether attitudes towards grades as an admission criteria is influenced by information about factors that influence grades, but are beyond individual control. In the following, I explain in more detail the design and instructions given to the participants. The complete survey is provided in the appendix.

3.1 Experiment

The experiment was conducted in Norwegian. This instructions are thus translated.

General introduction: *Imagine two people, person A and person B. They have the same educational program and school as their first choice in their application to upper secondary education. Imagine that there is only one seat left, and you are to decide which one of the two should receive that seat. You have the following alternatives: give the seat to person A, give the seat to person B, leave the seat empty or lottery draw. If you choose lottery draw we will ask you to distribute lottery tickets between the two. We are interested in your opinion, and thus there are no right or wrong answers.*

Treatment 1: Effort inequality *Person A has a grade average of 4,6, while person B has a grade average of 4.4.* ***Person A has spent more time on school work than person B.*** *What do you want to do?*

- *Give the seat to person A*
- *Give the seat to person B*
- *Lottery draw*
- *Leave the seat empty*

[If answer is "Lottery draw" the following additional question is displayed:]

You chose a lottery draw and we now would like to know how you want to distribute the lottery tickets. There is a total of 10 tickets. This means that if you give 10 tickets to person A then person A gets the seat for sure. If you give 0 tickets to person A it means that person B gets the seat for sure. If you give five tickets for

each both have the same chance of receiving the seat. How many seats do you want to give person A?

Treatment 2: Ability inequality The text in bold font in treatment 1 is replaced with:

Both have spent equally much time on school work, but person B is a slower learner than person A.

Treatment 3: Parental inequality The text in bold font in treatment 1 is replaced with:

Both have spent equally much time on school work, but person As has received help from the parents to get more out of the school work.

Treatment 4: Bad luck inequality The text in bold font in treatment 1 is replaced with:

Person B has been sick a lot and thus have missed much of the instruction.

3.2 The GPA difference

A GPA difference of 0.2 corresponds to about a 20% of the standard deviation in grades in Oslo municipality. It corresponds to about half of the estimated gender difference in grades (0.44) and the socioeconomic gradient of grades (0.3-0.39). Moreover, for the municipality of Oslo the median GPA was 4.58 in 2022 while the mean was 4.46. A reduction in grades of 0.2 points is equivalent from moving from the median to about the 40% percentile.

Each high school student receives about 14 grades. The grades are given in the interval 1 to 6 where 1 equals fail. A grade average of 4.64 corresponds to for example receiving 5 grades of 4 and 9 grades of 5. Reducing the GPA to 4.42 corresponds to reducing the GPA by 0.22, which for corresponds to 8 grades of 4 and 6 grades of 5. This means that improving your grade by 0.22 is equivalent to improving your grade one unit in three subjects, while holding the other grades constant.

3.3 Survey question

The participants also answer some survey questions. The survey questions are the same in all treatments and we will use the response to questions about locus of control.

3.4 Background questions

In addition the participants will answer the following set of background questions:

- Gender: (boy, girls, other/do not want to disclose
- Which grade point average do you anticipate to apply to upper secondary education with?
- Are you delivering an ordinary application to upper secondary education? (Yes, No (list of potential exemptions)

The last question is used to define the sample. We will only study ordinary applicants to upper secondary education. The other two are used for heterogeneity analysis.

3.5 Register data

- Social background
- Immigration status

These two are also used for heterogeneity analysis, more below.

4 Empirical strategy

4.1 Main outcome variable

The experiment is designed to study the attitudes towards different factors that are beyond individual control which may influence a persons school performance, and how these influence the attitudes towards merit based admission. The main focus is to compare how information about different sources of bad luck experienced by the low performer affect respondents' support for the high performer.

The support for the high performer is computed in the following manner: if the respondent chooses to assign the study seat to person A the outcome variable equals 1. If the respondent chooses to assign the study seat to person B or to leave the study seat open the outcome variable equals zero. If the respondent chooses a lottery then the outcome variable equals the number of tickets given to person A divided by 10. This means that the outcome variable can be interpreted as the probability that person A receives the study seat.

4.2 Hypotheses

4.2.1 Merit and luck

We first test whether introducing different luck components to person Bs performance reduces the support for person A. We apply a one-sided test for significance

since there is no reason to believe that these considerations could cause increased support for person A. We here test the effect on the support for the high performer, captured by the support for the individual with the highest grade, of introducing differences in ability (Treatment 2), parental support (Treatment 3) and brute luck (Treatment 4) relative to a situation where effort is the (main/articulated) source of performance difference.

Hypothesis 1 *Grade based admission has strong support in an environment where grade differences are attributed to differences in effort.*

Hypothesis 2 *The support for grade based admission is reduced when grade differences are attributed to factors beyond individual control.*

4.2.2 Different types of luck

Differences in ability, parental support and brute luck (captured by sickness absence) are different types of luck that influence an individuals performance. Ability, or talent, is related to effort as it influences the returns to effort. It is similar to productivity manipulations in experiments where agents receive different wages/multipliers for their exerted effort. Unlike in experiments it is a permanent trait so it will follow the individual into the future. Efficiency considerations may thus also come into play as it may say something about the individuals ability to succeed also in the future.

Parental support is also to some extent related to effort and influences the returns to effort. However, it is less a trait of the individual than ability. Moreover, it is less permanent in the sense that parental support in education might be expected to play less of a role when students get older. However, this is an empirical question for which we do not know the answer. However, these considerations leads to the following hypothesis.

Hypothesis 3 *Inequalities in merit are considered more fair when it is due to differences in ability than due to differences in parental support.*

Sickness absence is unrelated to the individual's choices and effort and thus more a form of brute luck. Moreover, it is likely to be temporary and not carry forward into the future. The information about the performance of an individual with a sickness absence is noisy as we don't know what the performance would have been in the absence of sickness. Based on this we state the following hypothesis:

Hypothesis 4 *Inequalities in merit are considered the least fair when an individuals effort has been constrained by factors beyond individual control.*

4.2.3 Heterogeneity

We will also study heterogeneity in inequality acceptance using the background data collected in the survey and from register data. We will focus on gender, own grades and socioeconomic background. Specifically we will test whether there are differences between the following groups along the same dimensions describe above:

- Education of parents: No parent with completed higher education and at least one parent with completed higher education.
- Low SES: Being in the bottom fifth of the income distribution and in the bottom fifth of the education distribution. See AlmÃ¥s et al (2015) for definition.
- Own grades: We will distinguish between people who have grades above person A (high performers), below person B (low performers) and between person A and person B (medium performers).
- Gender
- Locus of control: We will study how the choices and the treatment effects relates to the survey measure of locus of control

4.3 Specification and analysis

We here provide the main robust OLS regressions that will be used in the analysis.

4.3.1 Hypothesis 1-4

Hypothesis 1-4 will be tested by estimating the following regression equation:

$$s_i = \alpha + \delta_A A_i + \delta_{PS} PS_i + \delta_S S_i + \epsilon_i \tag{1}$$

where s_i is the support for the high performer by the respondent. A_i is an indicator value taking the value 1 if individual i had the ability treatment, PS_i is an indicator variable taking the value one if the individual had the parental support treatment, and S_i is an indicator variable taking the value one if the individual had the sickness treatment. T1- Effort is the reference group.

On the basis of equation 1 we test the hypotheses in the following way:

- Hypothesis 1
  - H:0 $\alpha \leq 0.5$
  - H:1 $\alpha > 0.5$
- Hypothesis 2:
  - $H0 : \delta_A = 0$ & $\delta_{PS} = 0$ & $\delta_S = 0$
  - $H1 : \delta_A < 0$ & $\delta_{PS} < 0$ & $\delta_S < 0$
- Hypothesis 3:
  - $H0 : \delta_A = \delta_{PS}$
  - $H1 : \delta_A < \delta_{PS}$
- Hypothesis 4:
  - $H0 : \delta_{PS} = \delta_S$ & $\delta_A = \delta_S$
  - $H1 : \delta_{PS} > \delta_S$ & $\delta_A > \delta_S$

4.3.2 The heterogeneity analysis

The heterogeneity analysis for gender will be conducted by estimating the following regression estimation:

$$s_i = \alpha + \alpha^F F_i + \delta_A A_i + \delta_A^F A_i F_i + \delta_{PS} PS_i + \delta_{PS}^F PS_i F_i + \delta_S S_i + \delta_S^F S_i F_i + \epsilon_i \quad (2)$$

where F_i is an indicator taking the value 1 if participant i is female. We will use corresponding regression equations for the other dimensions of heterogeneity. The formal statement of the hypotheses on heterogeneous effects correspond to Hypotheses 3-5.

In order to study whether the respondents choices in the experiment is related to their locus of control we will run the following regression.

$$l_i = \alpha + \alpha^e e_i + \delta_A^e A_i e_i + \delta_A A_i + \delta_{PS} PS_i + \delta_{PS}^e PS_i e_i + \delta_S S_i + \delta_S^e S_i e_i + \epsilon_i \quad (3)$$

where l_i is the measure for locus of control. On the basis of this regression we can for each of the treatments investigate whether there is a relationship between the locus of control and the support for grade based admission.

4.3.3 Balance and attrition

To test for balance, we will regress treatment on the main independent variables, both individually and together. With many variables tested, some of them are likely to be different and we will conduct an F-test of whether the control variables jointly predict treatment status.

Treatment are assigned in the survey at the entry to the question. We thus only have to worry about attrition during that question. We will check whether missing response to the question is correlated with treatment. If there are any statistically differences in attrition or non-response between treatment groups, we will follow the correction proposed by Lee (2009).

5 Exploratory analyses

We have chosen the fractional support for person A as the main outcome variable. However, alternative definitions could have been chosen. For instance, we could have chosen to study the share of respondents who give full support for person A, i.e., those that choose to give the study seat to person A for sure. This definition would give an idea of the distribution of individuals with different fairness ideals and thus could inform the results further. In the same spirit, we could also study the share who chooses to implement the lottery. These analyses might be conducted to shed light on the results from the main analysis.

Treatment T2, T3 and T4 describes factors that are beyond individual control. Drawing attention to such factors may influence people's perception of whether differences in grades are driven by differences in student effort rather than luck. As an exploratory analysis we will thus investigate whether there are any differences across treatments in their response to the questions: "My grades are a result of effort" and "My grades are a result of luck". Similarly, it might influence respondents perception of a grade-based admission system. After the treatment we ask the respondents whether they find it fair that grades determines who should get priority to a given school, and whether they find it fair that grades determines who gets priority to a given educational program.

6 Power calculation and discussion of null findings

We have four main hypotheses, but also four treatments. Because each of the hypotheses belong to different parts of the data we do not account for multiple hypothesis testing in these main hypotheses. Treatment is assigned at the individual level. We have recruited 34 classes, assuming 20 respondents in each class we have an endline sample of 680 individuals, 170 in each treatment. With 80 percent power and 5 percent significance and a one-tailed test we can detect an effect size of 0.27 standard deviations. The calculations does not account for the fact that we have baseline data with variables that may explain some of the variation in outcomes.

Some of our findings are likely to be null results. It is often difficult to judge whether such results are showing a meaningful lack of effect or whether they arise due to low power. To investigate if the effects are meaningful null findings we will conduct equivalence tests with two one-sided t-tests (TOST) and show how large positive and negative effects we can reject. The tests are one sided in equivalence testing as one tests whether effects are larger than a highest value and lower than a lowest value. In practice, the procedure is equivalent to presenting the bounds of a 90 percent confidence interval.

7 IRB approval and data protection

Norway does not have a system with IRB approval for research projects. The study was reviewed and approved by NIFUs data protection officer and received advice from Norwegian Centre for Research data. Informed concent for storing and combining data is given by the participants.

References

Bhuller, M., Mogstad, M., and Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4):993–1030.

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after california's proposition 209. *The Quarterly Journal of Economics*, 137(1):115–160.

Bleemer, Z. and Mehta, A. (2022). Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*, 14(2):1–22.

Canaan, S. and Mouganie, P. (2018). Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics*, 36(2):395–436.

Cohodes, S. R. and Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4):251–285.

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile. Technical report, National Bureau of Economic Research.

Kirkebøen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.