Essays in Applied Microeconomics

by

Roberto Carlos Asmat Belleza

Thesis
Submitted to the University of Warwick
for the Degree of
Doctor of Philosophy in Economics

Department of Economics

September 2021

Contents

List of Figures **i**

List of Tables **ii**

Acknowledgements **iv**

Declaration **v**

Abstract **vi**

1 Gender Differences in Judicial Decisions **2**

1.1 Introduction . . . 2
1.2 Overview of the Peruvian child support system . . . 7
1.3 Data . . . 8
1.3.1 Judicial Records . . . 8
1.3.2 Transparency Agency . . . 10
1.4 Empirical Analysis . . . 10
1.4.1 Identification condition . . . 10
1.4.2 Econometric specifications and results . . . 12
1.5 Robustness analysis . . . 13
1.5.1 Going beyond gender . . . 13
1.5.2 Sample selection bias . . . 14
1.6 Understanding the effect of incomplete information . . . 16
1.6.1 Conceptual Framework of Judicial Decision-making . . . 16
1.6.2 Methodology . . . 17
1.6.3 Econometric specifications and results . . . 19
1.6.4 Estimates of beliefs . . . 20
1.7 Conclusions . . . 24

2 Positive Feedback and Score Comparison **41**

2.1 Introduction . . . 41
2.2 Institutional Background and Data . . . 45
2.2.1 Background . . . 45
2.2.2 Data and Analysis Samples . . . 47
2.3 RD-based Analysis of Positive Feedback . . . 48
2.3.1 Research Design . . . 48
2.3.2 Validity of RD Design . . . 49

2.3.3 Impact on Future Participation 51
2.3.4 Gender Differences in Reaction to Feedback 53
2.4 Positive Feedback and Score Comparison 54
2.4.1 Descriptive Evidence 55
2.4.2 Exposure to Minimum and Maximum Scores 57
2.4.3 Exposure to High- and Low-performing Peers 58
2.5 Conclusions 60

3 Ordinal Rank Effects in Math Competitions 83
3.1 Introduction 83
3.2 Institutional Setting 87
3.2.1 Math Olympiads in Slovakia 87
3.2.2 School system in Slovakia 89
3.3 Data 90
3.3.1 Sample description 90
3.3.2 Ordinal Rank 92
3.3.3 Outcome variables 93
3.4 Empirical Analysis 93
3.4.1 Identification Strategy 93
3.4.2 Model Specification 94
3.4.3 Rank Effects on Outcomes 95
3.5 School Choices 96
3.5.1 Mathematical Training in Grammar Schools 97
3.5.2 Rank Effects on School Choices 99
3.6 Conclusions 100

Appendix 117
A.1 Randomisation of cases in the Judicial System (before) 117
A.2 Randomisation of cases in the Judicial System (after) 118
A.3 Olympiads Report - Example 1 119
A.4 Olympiads Report - Example 2 120

List of Figures

1.1 Kernel distributions of awards by judge's gender in formal cases 25
1.2 Kernel distributions of awards by judge's gender in informal cases 26
1.3 Kernel distribution of calibrated awards in informal cases 27
1.4 Kernel distribution of estimated judges' beliefs about the respondent's income 28
1.5 Kernel distribution of petitioner's claims 29

2.1 Histogram of Running Variable for RD Analysis 62
2.2 Histogram of Running Variable for RD Analysis by categories . 63
2.3 Predetermined Covariates by Points from Cutoff 64
2.4 Feedback Effect on Future Participation 65
2.5 Feedback Effect on Future Participation by categories 66
2.6 Feedback Effect on Future Participation by Gender 67
2.7 Feedback Effect on Future Participation by Maximum Score Achieved in Districts 68
2.8 Feedback Effect on Future Participation by Minimum Score Achieved in Districts 69

3.1 Math Olympiads Structure 101
3.2 Slovak School System and Math Olympiads 102
3.3 Variation in Rank 103
3.4 Subsequent Performance in Regular vs Grammar Schools 104

List of Tables

1.1 Sample characteristics . . . 30
1.2 Balance Table, Case Characteristics by Judge Gender (Hearing Settlement) . . . 31
1.3 Balance Table, Case Characteristics by Judge Gender (Litigation - Formal) . . . 32
1.4 Balance Table, Case Characteristics by Judge Gender (Litigation - Informal) . . . 33
1.5 Judge's gender effects on child support decisions . . . 34
1.6 Balance Table, Judges' Characteristics by Judge Gender . . . 35
1.7 Pooled OLS estimates with all judges' characteristics . . . 36
1.8 Marginal effects for settlement . . . 37
1.9 Estimates for settlement agreements . . . 38
1.10 Gender gap decomposition - pooled OLS estimates . . . 39
1.11 Pooled OLS estimates beliefs formation . . . 40

2.1 Sample Characteristics . . . 70
2.2 The Effect of Positive Feedback on Predetermined Characteristics of Competitors . . . 71
2.3 Effect of Positive Feedback on Subsequent Participation . . . 72
2.4 The Effect of Positive Feedback on Subsequent Participation by Gender . . . 73
2.5 Sample Characteristics - Upper Score Comparison . . . 74
2.6 Sample Characteristics - Lower Score Comparison . . . 75
2.7 Tests for Random Assignment of Maximum and Minimum Scores in District . . . 76
2.8 Positive Feedback and Score Comparison (Max and Min) . . . 77
2.9 Tests for Random Assignment of High- and Low-performing Peers (Definition 1) . . . 78
2.10 Tests for Random Assignment of High- and Low-performing Peers (Definition 2) . . . 79
2.11 Tests for Random Assignment of High- and Low-performing Peers (Definition 3) . . . 80
2.12 Positive Feedback and Score Comparisons (High- and Low-performing Peers) - district-by-year fixed effects . . . 81
2.13 Positive Feedback and Score Comparisons (High- and Low-performing Peers) - separate fixed effects . . . 82

3.1 Sample Characteristics . . . 105

3.2 Ordinal Rank Statistics 106
3.3 The Impact of Rank on Subsequent Participation 107
3.4 The Impact of Rank on Subsequent Performance 108
3.5 Grammar School Effects on Subsequent Performance 109
3.6 The Impact of Rank on School Choices 110

Acknowledgements

I would like to thank the following people for their support, without whom this thesis would not have been possible to write. First of all, I am indebted to my supervisors for their guidance and support. Thanks Roland Rathelot, for taking my research to new heights and encouraging me in bad times. Thanks Ben Lockwood, for helping me to develop logical thinking skills and your readiness to help me in difficult situations. I am also grateful to the faculty at the Economics Department whose comments benefited this thesis. In particular, thanks Manuel Bagues, for your insights that steered me through my research. I am also thankful to my peers who made the PhD years more enjoyable. In particular, Lajos Kossuth, Alessandro Castagnetti, Carlos Cardona and Daniel Diez.

And my biggest thanks to my parents and sisters for all the support you have shown me throughout this journey. Thanks Ma and Pa, for your hard work that gave me the opportunity to access to better education. Thanks Carol, Ysabel and Rosario, your academic achievements despite the socioeconomic barriers that surrounded us showed me I am capable to pursue my professional goals.

Declaration

This thesis is submitted to the University of Warwick in accordance with the requirements of the degree of Doctor of Philosophy in Economics. I declare that it has not been submitted for a degree at another university. Chapter 1 is co-authored with Lajos Kossuth (PhD candidate at Warwick Business School). I am the sole author for Chapters 2 and 3.

September 2021

Abstract

This thesis consists of three essays on Applied Microeconomics. It broadly deals with understanding how access to information can affect decision-making by using state-of-the-art causal analysis.

Chapter one studies how the lack of information about cases' characteristics affects differences in decision-making between female and male judges. This study uses administrative data on child support cases where the father is the respondent and the mother who has the child custody is the petitioner. By exploiting random assignment of cases to judges, this chapter reveals that female judges set lower awards than their male counterparts in child support cases. However, the gender-based difference is much lower when the income of the respondent is not observable to judges. By combining decisions made in cases where the income of the respondent is observable with decisions made in cases where it is not, the analysis shows that female judges estimate higher levels for the unknown income, which attenuates the gender-based differences in decision-making.

The other two chapters investigate how information given to competitors in district mathematical Olympiad affects their willingness to participate again and their subsequent performance. Chapter two evaluates whether giving positive feedback to competitors increases their subsequent participation in mathematical Olympiad. To establish causality, I exploit a score cutoff that determines the provision of positive feedback ("you are successful") and find that positive feedback positively affects subsequent participation in competitions. Interestingly, the positive feedback effect is weaker when recipients are surrounded by extremely talented competitors in their district but remains the same when surrounded by low-performing competitors. Chapter three investigates whether equally talented competitors who learn they are differently ranked in their districts participate more and perform better in the following year. By exploiting idiosyncratic variation in the score distribution across districts, I find that higher-ranked competitors are more likely to attend the

competition the following year and perform better. In exploring mechanisms, I investigate whether these rank effects are driven by school choices. I find positive but non-significant rank effect on the likelihood of switching from regular to selective schools.

Chapter 1

Gender Differences in Judicial Decisions

WITH LAJOS KOSSUTH

We compare decisions by female and male judges in child support trials where a judge decides on the father's income allocation to children. We investigate two types of cases: 1) when fathers have a formal job, their income is known to judges, and awards are set as a fraction of it, and 2) when fathers work in the informal sector, their income is unknown to judges, and awards are given as a fixed amount of money. By exploiting random assignment of cases to judges, we find that female judges set lower awards in both cases in comparison to male judges. However, the gender gap under incomplete information is around two thirds the size of that under complete information. In exploring mechanisms underlying this difference, we propose a simple framework to elicit judges' beliefs about the unknown income in such cases by using their judicial behaviour in cases where income is known. We find that female judges estimate that the unknown income is higher than male do. Thus, gender differences in estimated beliefs act as a countervailing force and explain the attenuation of judges' gender differences in decisions under incomplete information.

1.1 Introduction

The economics literature presents evidence on the factors affecting child support allocation when it is decided by a bargaining process between separated parents (Chiappori and Weiss, 2007). It is surprisingly less informative on how transfers from the non custodial parent to the child are determined when there is disagreement and they are adjudicated by a court. In this setting, the judiciary follows set legal guidelines for child support allocation with some room for discretion, leaving room for additional factors to influence the outcomes. In this paper, we explore one of these factors: the gender of judges. In particular, do male and female judges have different perspectives on a 'fair' allocation of child support?

Any attempt to map judges' preferences from observed judicial decisions faces an important empirical challenge: the judiciary is often not able to observe all relevant information to make a decision. Hence, these decisions may reflect both judges' preferences about the fair amount of child support to allocate but also their beliefs about the missing information. In particular, the vital piece of information in child support cases is the income of respondents, since it is used as a reference by judges to reach a verdict. When the respondent works in the informal economy, having complete knowledge about his income might be difficult, so judges must form beliefs about the unknown income before deciding how much child support to allocate. This issue is more salient in developing countries due to the large informal labour markets, which affects judicial decision-making (Sadka, Vera and Woodruff, 2018).

To address this, we combine judicial decisions in randomly assigned child support cases with complete and incomplete information about the income of respondents. In the former, since the respondent works in the formal economy (from now on, formal cases), the judge is able to observe his income, and decides on a child support amount that is expressed as a percentage of the respondent's income. In the latter, however, the respondent works in the informal economy (from now on, informal cases) and judges first need to form beliefs about his unknown income before deciding a child support amount, which is ultimately being awarded in absolute terms. Our sample therefore comprises judges that make decisions in both contexts, giving us a unique opportunity to inspect the degree to which incomplete information might play a role in shaping gender differences in the allocation of child support, but also in the elicitation of beliefs about the income of respondents as a potential mechanism.

We use the Peruvian judicial institutional setting for the following reasons. First, the features of the justice system allow us to identify a causal effect of gender differences in judicial decisions, since child support cases are randomly assigned to female and male judges. Second, due to the highly informal nature of the labour market, the distribution of cases across the formal and informal economies allows us to conduct the analysis described above. Third, the legal objective in child support cases is narrow and measurable: the award of a reasonable amount of money to meet the needs of the children. There are no other motives such as punishment or deterrence typically found in criminal cases. Finally, child support cases are heavily gender-coded. In Peru, women are generally expected to take care of children and men to provide income and physical protection. Indeed, more than 98% of the cases we analysed involve a mother suing a father.

Our database includes over 3,000 published child support cases in Lima, Peru, where the petitioner is always the mother and custodial parent, and the respondent is always the father. We extract information from two stages of a typical child support case. In the first stage, the settlement hearing, the parties are encouraged to negotiate and agree on a child support amount in order to avoid the expenses of proceeding to litigation. If the parties fail to settle, the case then proceeds to litigation. Here, judges have to decide a child support amount based on the evidence presented by the parties. Depending on the job type of the respondent, the ruling is a percentage of his income (formal job) or a fixed amount of money (informal job).

Our central finding is that, relative to their male counterparts, female judges allocate smaller amounts of child support per child. Moreover, we also find that these gender differences depend on whether the income of the father is observable. The gender-based gap is -6.8% in formal cases (0.25 standard deviations) and -5.9% in informal ones (0.16 standard deviations). These results are striking for two reasons. First, female judges seem to be more harsh towards the female petitioners, although this finding is in line with some evidence found in decisions about employment (Bagues and Steve-Volart, 2010) and academic evaluations (Bagues, Sylos-Labini and Zinovyeva, 2017). Second, results show that the gender-based gap found in formal cases - when there is complete information about the income of respondents - is 56% higher than that in informal ones.

To understand how the lack of information about the income of respondents attenuates the gender-based gap, we develop a simple framework of judicial decision-making in which incomplete information operates through beliefs. We interpret the verdict reached by a judge in a formal case as their revealed child support allocation preference. We then assume the same judge should exhibit the same preference in an informal case, the difference being the degree to which they would also need to form a belief about the income of the father. To illustrate, if the award given by a judge in a formal case is 30% of the respondent's income and the same judge awards S/. 300 in an informal case with observably similar respondents, we then infer that the judge must believe the respondent earns S/. 1,000. So, we use the estimated preference in the formal case, together with the child support amount allocated by that same judge in the informal case, to infer how they would form a belief about the respondent's income in that same informal case.

By restricting the sample to judges who face formal and informal cases and applying the framework described, we find a gender-based gap of around -4.7% in cases where income is unknown. This gender gap is composed of a gender gap in preferences for child support allocation of -12% and a gender gap in estimates of fathers' income of 7.33%. This means female judges - relative to their male counterparts - infer higher levels of income when information about it is incomplete or non-existent. In other words, even though female and male judges are exposed to similar sets of incomplete information (do to random allocation of cases to judges), female judges exhibit a higher reference point (estimate of the income of respondents) to decide on a 'fair' allocation of child support. Thus, gender differences in estimates act as a countervailing force and explain the attenuation of gender differences in decisions under incomplete information.

To gain a better understanding of why judges' estimates of fathers' income differ by gender, we explore the extent to which judges rely on the mother's claim (amount of money) as a signal about the unknown income of respondents. In informal cases, mothers state their claims arguing that this amount of money is a fair estimation of what the father is able to pay. It is worth noting that claims are recorded before the case is randomly assigned to a judge, and so are independent of the gender of judges. We study the relationship between beliefs and claims by using a standard Bayesian updating framework in which judges estimate the unknown income based on their priors and the signal (claim) sent by mothers. We find that female judges rely less on the claim when estimating the unknown income. This might suggest that one underlying explanation for the lack of homophily found in the reduced-form estimates is that female judges rely less on claims made by female petitioners than male judges.

Finally, we also consider how other characteristics of judges influence decision-making. Other explanations for the gender gap might be differences in age or work experience of judges. We collected data on age, work experience and self-reported wealth of judges from official sources. First, we check whether judges significantly differ across those characteristics and find that male and female judges are similar. Second, we include those characteristics as controls in our baseline regressions. We find that only for informal cases, judge's age and status have a significant negative and positive effect on the award, respectively. However, the gender gap remains statistically significant and we find no evidence of a change in the magnitude of the gap.

Our paper contributes to the large literature on the role of gender in judicial decision-making. Most of this research has focused on criminal cases (e.g. Gruhl, Spohn and Welch (1981); Coontz (2000); Collins and Moyer (2008)), discrimination cases (e.g. Farhang and Wawro (2004)) and a range of civil rights issues such as immigration appeals (see Gill, Kagan and Marouf (2015)) and issues affecting women[1] (e.g. Martin (1989); Peresie (2005); Boyd, Epstein and Martin (2010); Boyd (2016)), all in the context of Common Law. To the best of our knowledge, this study contributes to the unexplored branch of family law and child support cases, which has several advantages for the study of judges' gender differences in judicial decisions.

First, in many settings, judicial decisions are made by a panel of two or three judges (see Gill, Kagan and Marouf (2015) and Peresie (2005)) and the interaction between them poses the additional problem of how to disentangle the views of female and male judges. This is the main challenge to identification of judge gender effects in this literature. However, in our setting, a single judge decides child support cases so we can attribute differences to the gender of judges by exploiting random assignment. Second, the objective in child support cases is simple and narrow, as opposed to the typical issues studied in the literature (see Arnold, Dobbie and Yang (2018)). By isolating one legal objective, we can pin down preferences of judges on the single issue of income support for children. Third, as Kiser, Asher and McShane (2008) show, in most judicial instances, only 5% of cases reach litigation, so the interest should be in the settlement stage instead of the litigation stage. In child support cases, we find that 70% of cases that started the process reach litigation. Finally, in many instances there is female under-representation in courts, which poses an empirical threat when it comes to comparing decisions of male and female judges (Knepper (2018)). In stark contrast with other settings, in child support trials in Peru, 60% of judges are female.

This paper also contributes to a novel literature in attempting not only to detect outcome disparities, but also to study the reasoning and to learn

[1]The issues included in these studies are abortion, affirmative action, sex discrimination in employment and sexual harassment

about the decision-making process behind these results. For instance, Arnold, Dobbie and Yang (2018) test for racial bias in bail decisions by comparing misconduct rates of respondents for whom perceived benefits and costs of being released were equal for judges. In another recent study, Ash, Chen and Ornaghi (2020) look at gender attitudes to explain voting behaviour in gender-related cases in U.S. Circuit Courts. In the same spirit, our paper contributes to understanding gender differences in judicial decision-making by inspecting the role of incomplete information in shaping potential differences.

We structure this paper as follows. In Section 1.2, we describe the child support system in Peru. In Section 1.3, we review the data and construction of variables. In Section 1.4, we explain the random assignment of cases to judges, provide evidence in this regard, and then report the judge gender-based gaps. In Section 1.5, we show robustness checks and address sample selection concerns in our setting. In Section 1.6, we present a framework of judicial decision-making and discuss gender differences in estimates about the income of respondents as the main mechanism explaining results. Finally, the conclusion is in Section 1.7.

1.2 Overview of the Peruvian child support system

The goal in child support cases is for the judge to make an award (the specific monetary amount for child support the respondent will have to pay) based on the claims of both parties and all the available information. The criteria for setting an award is provided by the Peruvian Civil Procedure Law: a) the needs of the children who are trial matter; b) the respondent's income; and c) additional children the respondent must support. However, the way to balance these variables and how to determine the award is not stated explicitly and is at the discretion of each judge. The only firm rule is that the amount, in total, should not represent more than 60% of the respondent's income.

A crucial aspect of the legislation is that it is not necessary for a judge to know the respondent's income to set an award. About 70% of the workers in Peru work in the informal sector, and often a child support case involves adjudicating two opaque income streams. Therefore, judges' awards depend on the respondent's job status. First, in cases where the respondent has a formal job and his income is known to the judges, they set an award as a percentage of income. For the sake of simplicity, we define these cases as "formal". Second, in cases where the respondent works in the informal sector, his income is unknown to the judge, who sets an award as a fixed sum. We defined such cases as "informal". In these type of cases, a reference of the respondent's income is supposed to be made by judges, although they do not often report it. In this regard, the legislation states that the reference cannot be less than the legal minimum wage.

The process that a case goes as follows. Once a child support case is filed, it will first be revised by a court and subsequently admitted for trial. If admitted, the respondent is immediately notified and has up to five days to respond to the claim. If the respondent does not respond, he is declared a "rebel" and loses his right to present evidence that supports his position. After this step, a date is set for the first stage of the process under study: the settlement hearing. The judge's goal in this stage is to get both petitioner and respondent to settle, so they can avoid incurring in the monetary and time-consuming costs of litigating. If the petitioner and the respondent fail to settle on a specific amount for child support, the case proceeds to litigation where the judge will decide the award. Finally, if any or both parties disagree with the award, they could appeal the decision and proceed to a final stage.

1.3 Data

This paper uses data from two Peruvian administrative sources: virtual archives of judicial records ("Consulta de Expedientes Judiciales") and the national transparency agency ("La Contraloría").

1.3.1 Judicial Records

We use publicly available documents from the CEJ website relating to child support cases filed in the capital city of Lima during 2017 and 2018. CEJ provides all records of each action taken by the parties and the judge assigned in each case. For a given case, we collected documents corresponding to the two stages described in section 1.2: the settlement hearing (first stage) and the litigation (second stage). The final data set was built based on 3,015 child support cases from the website. In the following subsections, we describe how we extracted variables from both documents.

Settlement hearing: This document is signed by the judge assigned to the case and records the attendance of parties and their attorneys (names and IDs), and characteristics of children who are trial matter (full names and age).

As the main goal is to promote agreement between the parties, the following steps of the settlement process are conducted by the judge: i) the petitioner asks for an amount for child support, ii) the respondent either accepts or offers a different amount, iii) the judge suggests an amount, and iv) parties accept or reject the proposition. Unfortunately, most cases do not record this process and only show a no settlement/settlement result. We infer the judges' gender from their full names shown in the digital signature.

Litigation: This document is also signed by the judge assigned to the case[2] and contains the verdict and the judge's arguments. We extracted data from the 'case analysis' section[3] that contains the judge's analysis on the three criteria established by Peruvian law to determine an award. For the first criterion, the needs of the children who are trial matter, judges typically state their age, how much the mothers spend on them, and if they have any special needs such as health conditions. For the second criterion, the economic capacity of the respondent, the judge examines all his income sources, such as salary, businesses, and properties, should there be information about them. As explained in section 1.2, it is not necessary to thoroughly investigate the income of the respondent to provide an award according to the law, and this is especially relevant for cases where the respondent works in the informal sector. Finally, for the third criterion, the judge investigates whether the respondent has other dependent children to support[4].

Table 1.1 shows a summary of the data sets used in this study. Since only districts that have more than one court apply randomisation of cases to courts, we discarded cases from districts with only one court. At the settlement hearing stage, there are 2,371 cases in total. These cases were assigned among 149 judges, 59% of whom were women. 27.4% of the cases reached a agreement and did not proceed to the ligation stage. At the litigation stage, there are 1,736 cases in total. These cases were assigned among 153 cases, 61% of whom were women. In 22.7% of cases, the respondent had a formal job.

[2]We only found a judge in the litigation stage different than the one observed in the settlement hearing in a few cases. This is mostly explained by an abnormal delay between these two stages such that the judge leading the court changed.

[3]The complete analysis consists of three parts: legal framework, case analysis and verdict. In all cases, the legal framework section contains the same information about judicial principles followed and, therefore, there is no variation in this regard across cases.

[4]Descendants who are 18 years old or older, and other relatives such as parents or siblings are not considered the responsibility of the defendant although respondents often claim to have such responsibilities.

1.3.2 Transparency Agency

We supplement the data of judicial cases by further characteristics of judges from two publicly available sources. Given their important public responsibilities, judges are closely supervised by the Peruvian Transparency Agency to detect irregularities related to corruption. From CVs, we collected judges' ages, job position (principal judge or alternate judge) and years of experience as principal at the time of the study. The position of a judge (principal/alternate) depends on professional achievements and experience as a judge. Moreover, principal judges earn higher salaries and are held in higher esteem than alternate judges. From the second source (financial situation), we collected judges' wealth. It worth mentioning that this is self-report information, which should include savings at the financial system and valuation of their assets (mostly real estate).

1.4 Empirical Analysis

1.4.1 Identification condition

Identification is achieved by cases being randomly assigned to judges within a judicial district. According to the Peruvian Civil Law, child support cases are randomly assigned to courts within a judicial district[5]. Most judicial districts contain more than one court, and each court is led by only one judge. Thus, districts with more than one court will follow a random algorithm for assigning cases to courts. This ensures a fair distribution of caseload across courts and also prevents petitioners from targeting their cases to more favourable judges. As courts are led by one judge, court randomisation means that cases are randomly assigned to judges.

The randomisation process is conducted as follows. First, to sue for child support, the petitioner must attend a Peace Court in the judicial district corresponding to the geographical district where she lives. The forms to file a lawsuit are designed to be simple and accessible even to those who do not have the means to pay for legal services. The form is entered into the IT system in an office called 'Mesa de Partes'. The case is randomly assigned to one court office out of several within the judicial district. Finally, the court office

[5] Due to different population sizes of geographical districts, the judicial system sets "judicial districts" to aggregate small population-sized districts. For instance, two small population-sized geographical districts "X" and "Y" can be merged into the judicial district called "$X - Y$".

receives the lawsuit file and the process described in 2.1 starts.

We provide a screenshot of the randomisation step in the system as is written in its user guide[6]. This shows how the person in charge has to register the case into the system (see Appendixes A.1 and A.2). We highlight the fact that it is impossible for the officer to manipulate the assignment of the case to a court. This randomisation pipeline has been confirmed by two separate sources in interviews conducted in Peru in 2019[7].

To corroborate that cases are indeed randomly assigned to judges in our data set, we conduct balancing tests. Table 1.2 presents the balance check of cases characteristics observed at the settlement hearings. For each group of judges, it presents the means and standard errors of the variables used for the analysis. The balance check is determined by the p-values of the differences-in-means two-tailed t-tests shown in the last column. It is important to note that these calculations only contemplate judicial offices in which there is at least one male and one female judge, as randomisation of cases to judges could be conducted. A statistically significant p-value suggests that there is enough statistical evidence to reject the null hypothesis of balance. As can be observed in Table 1.2, there is no evidence of imbalance in any of the variables observed. Therefore, we have no reason to doubt that settlement hearings were randomised across female and male judges.

As our main analysis is based on cases that reached the litigation stage, we also check whether cases' characteristics are balanced across female and male judges at the litigation stage. Table 1.3 and Table 1.4 present balance checks for explanatory variables in the litigation stage for formal and informal cases, respectively. Analogously to Table 1.2 they present the means and standard errors of the relevant variables used for the analysis for male judges and female judges. The last column shows the p-values of the differences-in-means two-tailed t-tests. Again, these calculations only contemplate judicial offices in which there is at least one male and one female judge. In formal cases, Table 1.3 shows no imbalance except for one variable (with 90% of confidence): the number of children outside of trial that the respondent has to support. For informal cases, Table 1.4 shows no imbalance in all cases' characteristics.

All these pieces taken together constitute robust evidence that cases were randomised and the registration of cases followed the user guide described above.

[6]The user guide is available here (click to link).

[7]We interviewed the assistant of a judge who works in one of the courts in our study and a lawyer who had served as an attorney in child support cases.

1.4.2 Econometric specifications and results

Before we present the estimation of gender differences in decision-making, we provide a graphical analysis of decisions by gender of judges. Figures 1.1 and 1.2 show kernel distributions of child support allocations given by male and female judges in formal and informal cases. In both cases it can be clearly seen that the mean of the distribution for female judges is lower than that for male judges, indicating that female judges award lower amounts of child support on average. Moreover, the differences between female and male judges seem to be larger in formal cases than in informal cases. We estimate the judge gender gap in formal and informal cases by exploiting random assignment of cases to judges as follows:

$$log(\frac{\alpha_{ij}}{N_i^T}) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \epsilon_{ij}, \quad i \in F \quad (1.1)$$

$$log(\frac{A_{ij}}{N_i^T}) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \epsilon_{ij}, \quad i \in I \quad (1.2)$$

In equation (1.1), α_{ij} is the award (as percentage of the respondent's income) in formal case $i \in F$ assigned to judge j. $Female_{j(i)}$ is an indicator variable for whether formal case $i \in F$ was assigned to a female judge j. N_i^T and N_i^{-T} denote the number of children involved in formal case $i \in F$ and the additional number of children the respondent in formal case $i \in F$ needs to support, respectively. In equation (1.2), A_{ij} is the award (as a fixed amount of money) in informal case $i \in I$ assigned to judge j. $Female_{j(i)}$, N_i^T and N_i^{-T} are analogous variables for informal case $i \in I$. Finally, both equations include district γ_d and year γ_t fixed effects γ_d. All standard errors are clustered at the judge level. The main coefficient of interest is β_1 which estimates a semi-elasticity: the percentage change in the award when the case is assigned to a female judge relative to a male judge.

Table 1.5 contains the results of the pooled OLS regressions for the formal and informal cases. Columns (1) and (2) show that the gender-based gap is -6.8% in the formal cases, and -5.9% in the informal cases. For the formal cases, this means that female judges set an award per child that is on average 6.8% lower than that of the male judges. Looking at the other explanatory variables provides additional depth in understanding the judicial decision-making process in child support cases. Both the number of children included in the trial and the number of additional children the respondent has to support are negatively associated with the award in both formal and informal cases, since the maximum award a judge can give is 60% of the respondent's income. The coefficient associated with the number of children in trial is bigger when compared to the additional children the respondent needs to support in both formal and informal cases.

While the estimates of the gender gaps are easily interpreted (semi-elasticity), the limitation is that they cannot be comparable between formal and informal cases. Indeed, while the judge gender differences in formal cases correspond to differences in ratios (% of income awarded), in informal cases they correspond to differences in levels (amounts of money). To make them comparable we re-estimate equations (1.1) and (1.2) by standardising the awards given in formal and informal cases. Table 1.5 shows the results in columns (3) and (4). When a formal case is assigned to a female judge, the allocated child support amount per child is -0.25 standard deviations relative to when it is assigned to a male judge. The analogous figure for informal cases is -0.16 standard deviations. This means that the female judge effect is 56% stronger in formal cases than in informal cases. This striking result raises the question of how incomplete information has this attenuation effect in judge gender differences in child support decisions. We develop a simple framework to address the role of incomplete information in decision-making in section 1.6.

1.5 Robustness analysis

1.5.1 Going beyond gender

We have provided evidence of a gender-based gap when judges make child support rulings. 'Gender' might not be the only story behind these results, however. For instance, a female judge who is 60 years old at the time of trial might have views about what she considers a 'fair' allocation of child support, not because of her being female, but because she belongs to a generation with distinct social norms.

Thus, we check whether disparities are attributable to judges' characteristics other than their gender. As described in section 1.3.2, we collected data on their ages, job status (principal or alternate), years of experience as principal and self-reported wealth. First, we inspect whether male and female judges differ along these characteristics. Second, we include all judges' characteristics in our baseline regressions.

Table 1.6 presents a balance analysis for characteristics of judges by their gender. This test shows that female and male judges do not differ in other characteristics beyond gender except for their job status: 49.2% of male judges work as principal judges while for female judges it is 35.2%.

Table 1.7 shows the estimates of equations (1.4) and (1.5) with all judges' characteristics as control variables in addition to the variables used in Table 1.5. It can be seen that the gender effect prevails after including these controls. For formal cases, none of the other judge's characteristics are statistically associated with awards given by judges. For informal cases, judge's age, job status, and wealth have a significant effect on awards. However, there is no evidence that the magnitude of the gender gap changes after including them.

1.5.2 Sample selection bias

In this section we investigate whether the litigation stage in child support cases is suitable for detecting gender-based differences in judicial decisions. As Knepper (2018) pointed out, to detect judge gender-based differences in decisions, one must also inspect judge gender–based settlement rates. Indeed, the prior literature has focused on the trial stage and has ignored the fact that, optimistically, cases reach trial 20% of the time (Kiser, Asher and McShane, 2008). This could make the trial stage unsuitable for detecting gender-based differences if the judge's gender influences the likelihood of settlement and creates a sample selection bias in the trial stage, as Knepper (2018) finds in workplace sex discrimination cases.

Thus, we start with a hypothesis that the judge's gender has an effect on the probability of settlement. Exploiting random assignment of cases, we test this hypothesis as follows:

$$Pr(Settle_{ij} = 1) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 Formal_i + \gamma_d + \gamma_t + \epsilon_{ij} \quad (1.3)$$

Where $Settle_{ij}$ is an indicator variable for whether case i assigned to judge j settles or avoids litigation. $Female_{j(i)}$ is an indicator variable for whether case i was assigned to a female judge j. We control for N_i^T, the number of children involved in case i and $Formal_i$, an indicator variable for whether the respondent has a formal job. Regarding the latter, for cases that do not settle, the type of respondent's job is captured in litigation. While for cases that settle, we infer the type of job by the format that the agreed amount is expressed in[8]. Finally, γ_d and γ_t are district and year fixed effects, respectively.

Table 1.8 shows the marginal effects of regression (3). We find no gender effect on the likelihood of settlement. Moreover, in stark contrast with figures from discrimination cases suggested by Knepper (2018) where only 5% of cases reach litigation, in our data-set over 70% of cases failed to settle and proceeded to litigation. Interestingly, Table 1.8 shows that when the respondent has a formal job the likelihood of settlement decreases. Although the evidence indicates that the judge's gender does not affect the likelihood of settlement (extensive margin), this raises a concern about the potential effect of the judge's gender on the level of agreement (intensive margin) for cases that settle.

In light of that result, we also test the impact of the judge's gender on the level of agreement for cases that settle as follows:

$$Agreement_{ij} = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 Formal_i + \gamma_d + \gamma_t + \epsilon_{ij} \quad (1.4)$$

Where $Agreement_{ij}$ is the level of agreement (either in fixed amount of money or in percentage terms) for case i assigned to judge j that settles in the settlement hearing. $Female_{j(i)}$ is an indicator variable for whether case i was assigned to a female judge j, N_i^T is the number of children involved in case i and $Formal_i$ is an indicator variable for whether the respondent has a formal job. Finally, γ_d and γ_t are district and year fixed effects, respectively.

Table 1.9 shows the estimates of equation (1.4). Again, we do not find a significant effect of the judge's gender on the level of agreement for cases that settle. This means that there is no evidence that cases would self select into litigation depending on the judge's gender. Thus, we provide strong evidence that the litigation stage in child support cases is suitable for detecting gender-based differences.

[8] If the agreement is in percentage of income, we assume the respondent has a formal job, whereas when the agreement is a fixed amount of money, we assume he has an informal job.

1.6 Understanding the effect of incomplete information

The main result from the previous section is that the gender gap in formal cases is around 56% bigger than in informal cases. We propose a simple model of how incomplete information shrinks gender differences in child support decisions, and emphasise beliefs about income as the mechanism. We elicit these beliefs based on the methodological premise that the pool of judges who are making decisions in informal cases are revealing their preferences in formal cases during the period of analysis.

1.6.1 Conceptual Framework of Judicial Decision-making

For the sake of simplicity, let us assume a situation in which one female judge and one male judge have to make their decisions in trials where there is only one child involved, and where the respondent supports no other children. A case i is randomly assigned to a judge with gender $g = m, f$ who sets an award to be paid by the respondent. There are two types of cases, as mentioned in the previous section: the formal case $i \in F$ and the informal case $i \in I$.

In a formal case, a judge with gender g observes y_i - the respondent's income in the formal case $i \in F$ - and awards a monthly percentage deduction from the respondent's salary $\alpha^g \in (0, 0.6]$. Note that we allow α^g to vary between gender g and that it is constant within gender g since it does not depend on any case-specific characteristic. Let us reiterate that, legally, the judge only needs three pieces of information to decide on the case: the number of children who are trial matter, the respondent's income, and any additional children the respondent has to support. Given that the judge is able to observe y_i and that, by assumption, trials only involve one child and no other children to support, α^g is the fraction that the judge considers fair to deduct from the respondent's salary to support one child who is trial matter. Thus, the fraction of income per child is our outcome of interest in formal cases and what we define as the judge's allocation preferences.

In an informal case, on the other hand, a judge with gender g does not observe y_i - the respondent's income in the informal case $i \in I$ - and this time sets an award $A_i^g, i \in I$ to be paid monthly by the respondent. To do so, the judge forms a belief about the respondent's income $b_i^g, i \in I$ and sets a percentage $\alpha^g, i \in I$ to be deducted from it. It is worth noting that neither b_i^g nor α^g are observed in $i \in I$. We only observe the given award A_i^g. Thus, a judge with gender g will choose an award A_i^g in an informal case $i \in I$ as follows:

$$A_i^g = \alpha^g b_i^g \quad \forall i \in I \tag{1.5}$$

Again, the main outcome of interest is the award per child in an informal case. This is implicit since we assume there is only one child involved in the trial and no other children to support. Thus, under this framework, a judge with gender g grants an award that is constituted by their allocation preferences α^g and their beliefs about the respondent's income b_i^g.

This framework allows for interpreting the gender-based gaps in judicial decision making both in formal and informal cases. In the formal cases, the gap is straightforward to calculate since the award α^g only varies between gender g and does not depend on case-specific characteristics:

$$GAP_{formal} = \frac{\alpha^f}{\alpha^m}, \quad i \in F \tag{1.6}$$

If the gap in equation (1.6) were less than 1, it would mean that female judges have preferences for lower shares of income to be allocated to a child than their male counterparts in formal cases. As can be seen, this gap is entirely driven by differences in allocation preferences between genders.

In the informal cases, however, the gender-based gap would take the following expression:

$$GAP_{informal} = (\frac{\alpha^f}{\alpha^m})(\frac{\bar{b}^f}{\bar{b}^m}), \quad i \in I \tag{1.7}$$

If the gap in equation (1.7) were less than 1, it would mean that female judges are more lenient than their male counterparts towards the respondents in informal cases. However, the gap encompasses two different gaps: the gap in allocation preferences and the gap in beliefs about the respondents' income. We use this framework to interpret the results shown in section 1.4. Moreover, this framework is the starting point in understanding how judges' gender-based differences in beliefs might expand or shrink the gender-based gap in allocation preferences under some assumptions, as we explain in the next section.

1.6.2 Methodology

Intuitively, this methodology uses the revealed allocation preferences α_{ij} in the formal cases $i \in F$ assigned to judge j to infer judge's j beliefs about the income of respondents in informal cases $i \in I$. To illustrate, if the award given by a judge in a formal case represents 30% of the respondent's income and awards S/. 300 in a similar but informal case, we can then infer that the judge believes the respondent earns S/. 1,000 by assuming that the judge should maintain, *ceteris paribus*, the same allocation preference (30%). Thus, the award in informal cases breaks down into allocation preferences and beliefs about the respondent's income and we inspect the role of each of these factors in determining the gender-based gap.

Let us remember the gender-based gap expression in informal cases under the simple framework we developed in section 1.6.1:

$$GAP_{informal} = (\frac{\alpha^f}{\alpha^m})(\frac{\bar{b}^f}{\bar{b}^m}), \quad i \in I$$

Equation (1.7) computes the gender-based gap in informal cases I based on the given awards A_i^g which are observable. However, the factors of that decision, b_i^g and α_i^g, are unobserved by the researcher. If those variables were observable, we could inspect the role of uncertainty in shaping the gap in informal cases I, by taking log of equation (1.7):

$$log(GAP_{informal}) = log(\frac{\alpha^f}{\alpha^m}) + log(\frac{\bar{b}^f}{\bar{b}^m}) \tag{1.8}$$

If the gap in allocation preferences were less than 1 but the gap in beliefs more than 1, then we could conclude that incomplete information would attenuate the gap in allocation preferences. To conduct such an analysis, we propose the next methodology to calibrate α_i^g in order to estimate the parameter b_i^g.

We first calibrate the allocation preferences from the decisions made by judges in formal cases $i \in F$ as follows:

$$\frac{\alpha_{ij}}{N_i^T} = \mu_j + \beta' X_{ij} + \epsilon_{ij}, \quad i \in F \tag{1.9}$$

Where α_{ij} is the award given by judge j in formal case $i \in F$; X_{ij} is a vector that contains the main criteria for award-giving in child support cases (number of children in trial and additional children in need of support by the respondent), and μ_j is the judge fixed-effect.

Second, we infer how those same judges would have decided the award $\tilde{\alpha}_{ij}$ (as a percentage of income to be deducted) in informal cases. By using the coefficients from estimating equation (1.9), we predict the allocation preferences

$\tilde{\alpha}_{ij}$ in informal cases as follows:

$$\tilde{\alpha}_{ij} = (\hat{\mu}_j + \hat{\beta} X_{ij}) N_i^T, \quad i \in I \tag{1.10}$$

Where $\tilde{\alpha}_{ij}$ is the calibrated award made by judge j in informal case $i \in I$ as a percentage of the respondent's income; vector X_{ij} contains the same set of variables as in equation (1.9) but for informal case $i \in I$; $\hat{\beta}$ is the vector of coefficients estimated in equation (1.9), and $\hat{\mu}_j$ is the estimated judge fixed effect also taken from equation (1.9). Note that the total estimated award is the multiplication of the calibrated award per child and the number of children involved in the informal trial. Figure 1.3 shows the kernel densities of the calibrated awards in informal cases by judge's gender. It can be seen that awards vary from 0.1 to 0.5 in general and that distributions seem to have the same variance but not the same mean: the first moment for male judges might be higher than for female judges.

The third and last step is to estimate the judge's belief about the respondent's income b_i in informal case $i \in I$ by combining the calibrated award $\tilde{\alpha}_{ij}$ (as percentage of income) and the observed award A_{ij} (fixed amount of money) in informal case $i \in I$:

$$\hat{b}_{ij} = \frac{A_{ij}}{\tilde{\alpha}_{ij}}, \quad i \in I \tag{1.11}$$

Figure 1.4 shows the kernel distributions of estimates by judge's gender. As opposed to Figure 1.3, the mean and the variance of distribution of estimates seem to vary by judge's gender. Note that in all these estimations we are not interested in the effect that the judge's gender has on the awards: the judge's time-invariant characteristics (such as gender) are captured by the judge fixed effect μ_j. Instead our focus is on modelling the award in formal cases to predict how judges would have awarded a percentage deduction in informal cases.

1.6.3 Econometric specifications and results

Given the inputs provided by the expressions (1.10) and (1.11), we decompose the awards in informal cases A_{ij} into allocation preferences $\tilde{\alpha}_{ij}$ and beliefs $\hat{b}_{ij}$. Then we calculate the gender-based gap in both dimensions to measure the relative contribution of both sources of variation to the total gap in awards under incomplete information.

$$log(\frac{A_{ij}}{N_i^T}) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \epsilon_{ij}, \quad i \in I \quad (1.12)$$

$$log(\frac{\tilde{\alpha}_{ij}}{N_i^T}) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \epsilon_{ij}, \quad i \in I \quad (1.13)$$

$$log(\hat{b}_{ij}) = \beta_0 + \beta_1 Female_{j(i)} + \beta_2 N_i^T + \beta_3 N_i^{-T} + \gamma_d + \gamma_t + \epsilon_{ij}, \quad i \in I \quad (1.14)$$

The main coefficient of interest β_1 in equation (1.13) estimates the difference in the average allocation preference displayed by female versus male judges, while the main coefficient of interest β_1 in equation (1.14) estimates the difference in the average estimate of the respondents' income of female versus male judges for observably similar respondents. The results are in Table 1.10.

As can be seen in Column 1, the observed gap suggests that female judges are 4.7% more lenient towards respondents. Interestingly, however, the gap is bigger if we only take into account allocation preferences (12%, as shown in Column 2). This suggests that incomplete information about the respondent's income attenuates the gender gap in allocation preferences. Indeed, female and male judges respond differently to incomplete information: the former (compared to the latter) estimate that the income of respondents is 7.3% higher on average.

These estimates, moreover, are not driven by differences in specific case characteristics between genders. As it was shown in the balance tests (section 1.4.1), these are balanced between male and female judges. In particular, we look at the claim of petitioners as a possible source of variation. Figure 1.5 presents the kernel distributions of estimates by the judge's gender which show that the first and second moments do not differ by the judges' gender. In the next section, we develop a simple framework to investigate factors that could drive this result.

1.6.4 Estimates of beliefs

In this section, we propose a simple framework to estimate why male and female judges form different beliefs about the income of respondents. The starting point is based on the fact that randomisation of cases ensures that male and female judges decide on cases with the same set of (incomplete) information about cases' characteristics when respondents work in the informal sector. Therefore, it must be the case that male and female judges process the signals of the unknown income of respondents differently. In particular, we focus on the claim (the amount of money requested from the respondent by the petitioner) which is the main signal given by petitioners to judges on this regard. In fact, when setting their claims, petitioners argue that the amount asked corresponds to the level of income of respondents. Thus, we investigate to what extent male and female judges rely on this piece of information to estimate the unknown income.

This framework is based on the standard Bayesian updating model in which judges form beliefs about the respondent's unknown income Y based on a prior belief and a signal (the petitioner's claim amount C). However, we introduce one important feature to this simple model: the fact that petitioners exaggerate their claims to signal that the respondent has higher disposable income in order to increase the chances of obtaining higher awards. It is important to note that this signal, by definition, cannot depend on the gender of judges since it is set by the petitioner before the randomisation of the case. Thus, to create different income predictions based on the gender of judges, we allow the prior of judges about the unknown income to depend on their gender.

The model

We assume that the judge with gender g does not know the true respondent's income $y = ln(Y)$ but has a prior y_0^g and a fixed variance $(\sigma_0^g)^2$ of the prior about y:

$$y = y_0^g + \sigma_0^g \delta \quad \delta \sim N(0, 1) \tag{1.15}$$

While the petitioner's claim amount $c = ln(C)$ is an upward biased signal of the unknown income in the following fashion:

$$c = By + \sigma\epsilon \quad \epsilon \sim N(0, 1), B > 1 \tag{1.16}$$

Since the randomisation of cases occurs after the respondent makes the claim, the exaggeration rate B cannot depend on the gender of judges. Also, note that, as a simplification, the exaggeration rate B is constant and must be more than 1 to reflect the fact that the signal is an inflated version of the true income. Since the judge knows the claim has an exaggeration rate B, the judge cares about the deflated signal $\tilde{c} = \frac{c}{B}$. Thus, a more intuitive way of writing down equation (1.16) is:

$$\tilde{c} = y + \frac{\sigma}{B}\epsilon \quad \epsilon \sim N(0,1), B > 1 \tag{1.17}$$

Since $(y, \tilde{c})$ is distributed according to a bivariate Gaussian distribution, we pin down the judge's belief formation (posterior) based on the prior mean and the petitioner's deflated signal of the unknown income as follows:

$$E(y|\tilde{c}) = E(y) + \frac{Cov(y,\tilde{c})}{Var(\tilde{c})}(\tilde{c} - E(\tilde{c})) \tag{1.18}$$

Working with equations (1.15) and (1.17) to find the elements of equation (1.18):

$$E(y) = y_0^g, E(\tilde{c}) = y_0^g, Var(\tilde{c}) = (\sigma_0^g)^2 + \frac{\sigma^2}{B^2}, Cov(y,c) = (\sigma_0^g)^2 \tag{1.19}$$

Plugging these results in equation (1.18):

$$E(y|\tilde{c}) = y_0^g + \frac{(\sigma_0^g)^2}{(\sigma_0^g)^2 + \frac{\sigma^2}{B^2}}(\tilde{c} - y_0^g) \tag{1.20}$$

Equation (1.20) shows that if the deflated signal $\tilde{c}$ exceeds the judge's prior y_0^g, the judge's guess is adjusted upwards. Conversely, if the deflated claim $\tilde{c}$ is lower than the judge's prior y_0^g, the judge's guess is adjusted downwards. In other words, given the judge's prior y_0^g, the updating direction depends on whether the claim c exceeds or is less than the threshold By_0^g.

Estimation and results

To estimate parameters of this model, we replace $\tilde{c}$ in terms of c in equation (1.20) and we obtain the following:

$$E(y|c) = \theta y_0^g + \omega c \tag{1.21}$$

where $\theta = \frac{\sigma^2}{(\sigma_0^g)^2 B^2 + \sigma^2}$ and $\omega = \frac{(\sigma_0^g)^2 B}{(\sigma_0^g)^2 B^2 + \sigma^2}$.

In these expressions, the weights θ and ω on y_0^g and c, respectively, can be interpreted in terms of the variance (or precision) of the judge's prior about y, relative to the variance (or precision) of the petitioner's signal about y. For instance, the more accurate or precise the petitioner's signal (i.e. the lower is σ), the greater is ω. However, it is important to note that the exaggeration rate affects these weights through σ_0^g: B expands the negative effect of σ_0^g on θ and shrinks the positive effect of σ_0^g on ω.

Since the parameter σ allows for some noise in the claim as a signal of income and cannot depend on the gender of judges, we set it equal to 1 as a simplification and focus on the comparison of y_0^g and σ_0^g across gender given a constant exaggeration rate B. Thus, we estimate equation (1.21) separately for male and female judges as follows:

$$\ln(\hat{b}_{ij}) = \beta_1 + \beta_2 \ln(C_i) + \gamma_d + \gamma_t + \epsilon_{ij} \tag{1.22}$$

Where $\hat{b}_{ij}$ is the belief set by judge j about the income of respondent-case i estimated from equation (1.11), C_i is the claim amount made by the petitioner in case i. In addition, both equations include district γ_d and year γ_t fixed effects. Thus, we interpret $\hat{\beta}_1$ and $\hat{\beta}_2$ in equation (1.22) as θy_0^g and ω from equation (1.21), respectively. By assuming that $\sigma = 1$, we recover the parameters of interest σ_0^g and y_0^g:

$$y_0^g = \frac{\hat{\beta}_1 B}{B - \hat{\beta}_2 B^2} \tag{1.23}$$

$$(\sigma_0^g)^2 = \frac{\hat{\beta}_2}{B - \hat{\beta}_2 B^2} \tag{1.24}$$

Table 1.11 shows the estimation of equation (1.22) for each gender separately. It can be seen that the estimates differ by the gender of judges. Replacing $\hat{\beta}_1$ and $\hat{\beta}_2$ for each gender in the previous expressions, it can be shown that $y_0^m > y_0^f$ and $\sigma_0^m > \sigma_0^f$ for any value of $B \in (1, 3)$.[9] These results suggest that, although female judges have a lower prior mean, the associated weight on the prior mean is higher because of the higher precision (lower σ_0^g) of the prior mean in comparison to male judges. Regarding the weight on the signal, female judges rely less on the petitioner's claim because the precision of the signal relative to the precision of prior mean, $\frac{\sigma}{\sigma_0^g}$, is lower for female judges (given the higher precision of the prior mean) which decreases the weight on the signal in comparison to male judges.

[9] An exaggeration rate B larger than 3 would generate negative values of mean and variance priors of judges.

1.7 Conclusions

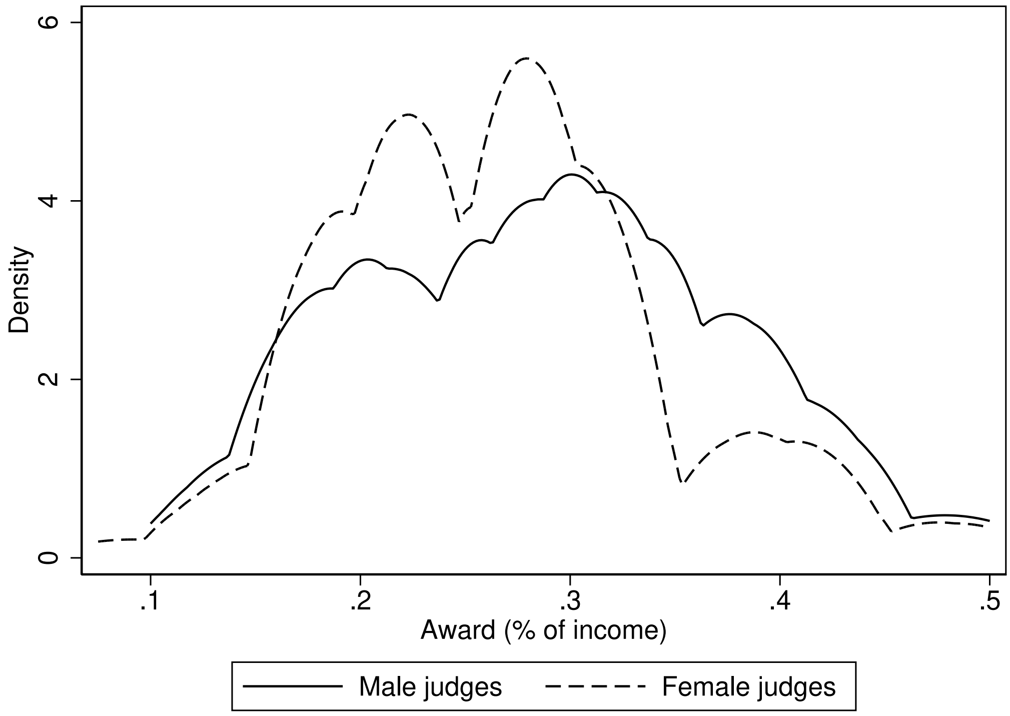
In this article, we inspect whether there are gender differences when child support is decided by a court rather than by negotiation between the parents. By exploiting random assignment of cases to judges, we find that female judges decide on a lower allocation of child support than male judges do in formal and informal cases. Moreover, we find that the effect of assigning a female judge to a formal case is 56% stronger than for an informal case.

We use a simple model of incomplete information. We assume that a judge has the same preferences for child support allocation regardless of whether a case is formal or informal. By estimating this component when the judge makes decisions in formal cases, we are able to calibrate their preferences in informal cases and infer the degree to which the amount of child support they allocate in the latter is influenced by their beliefs about the income of respondents. We find that, relative to male judges, female judges infer that the respondent has higher levels of income when they cannot observe it during trial, explaining why the gender gap in informal cases is smaller than in formal cases. By using a simple Bayesian updating framework in which judges form beliefs about the unknown income based on their priors and the signal (claim) sent by the petitioner, we provide a possible explanation for this fact: data shows that female judges rely less on the signal sent by the petitioner and put more weight on their priors. These findings highlight the fact that information asymmetries might play a role in influencing the outcomes of different types of judicial settings. For instance, could the lack of information explain racial disparities in the outcomes of criminal cases?

Finally, the evidence found in this paper has vital policy implications. There is evidence that parents transport less economic resources after parental separation (Bjorklund & Sundstrom 2006). For example, since the father has reduced access to the child, he has less incentives to provide resources. Further, if the mother remarries, the father has fewer incentives to support his child because part of the transfer spills over to the new husband (Chiapori 2007). Hence, child support allocation is not a trivial matter, so a discussion about the predictability of the judicial system in these type of cases is necessary, given that verdicts depend so much on variables such as the gender of the judge or the lack of information during trial. A potential solution to reduce discretion in child support cases could be for judges to rely on benchmarks based on, for instance, the type of respondent's job or the cost of living of the district where the child resides. This is an important issue for further research.

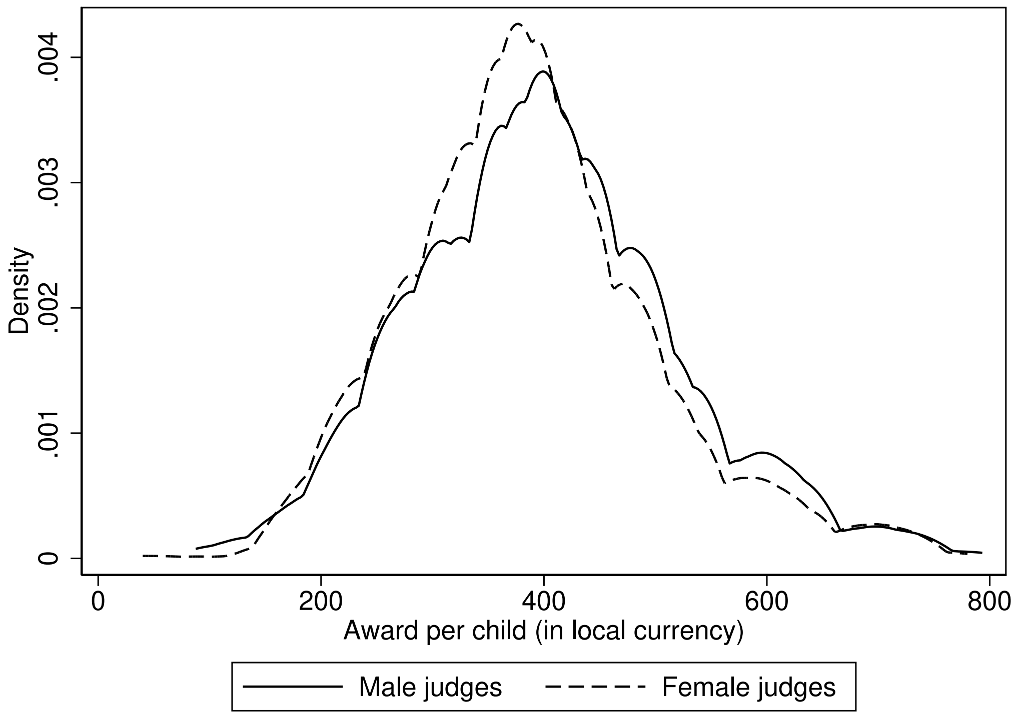
Figures

Figure 1.1: Kernel distributions of awards by judge's gender in formal cases



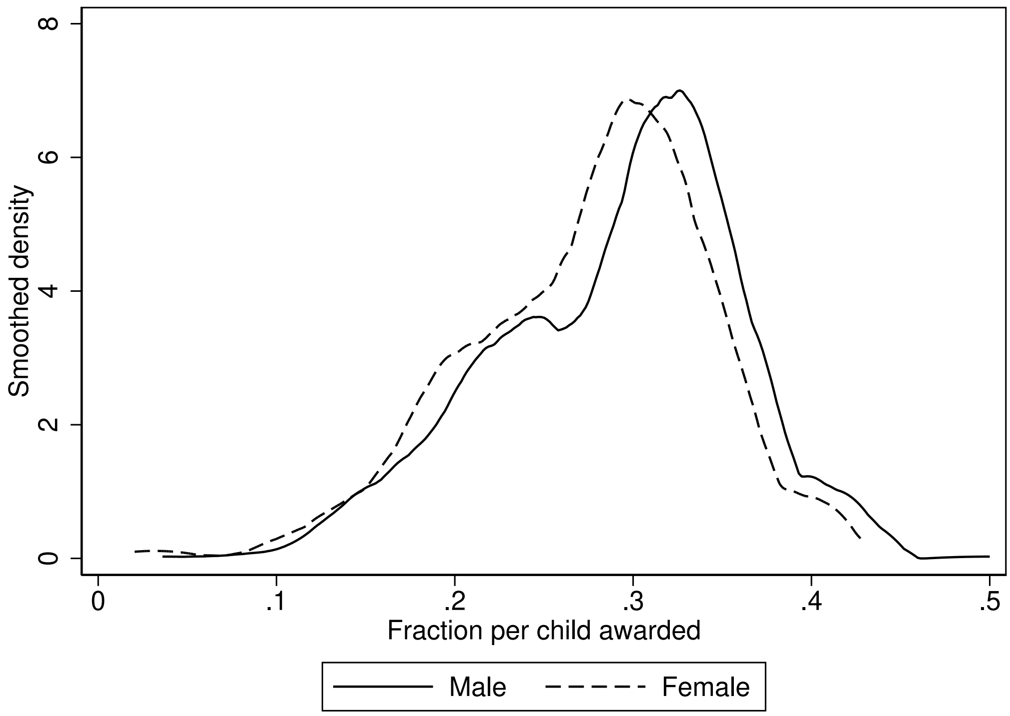
Notes: Lines show the kernel densities of awards (as a percentage of respondent's income) set by male and female judges in formal cases.

Figure 1.2: Kernel distributions of awards by judge's gender in informal cases



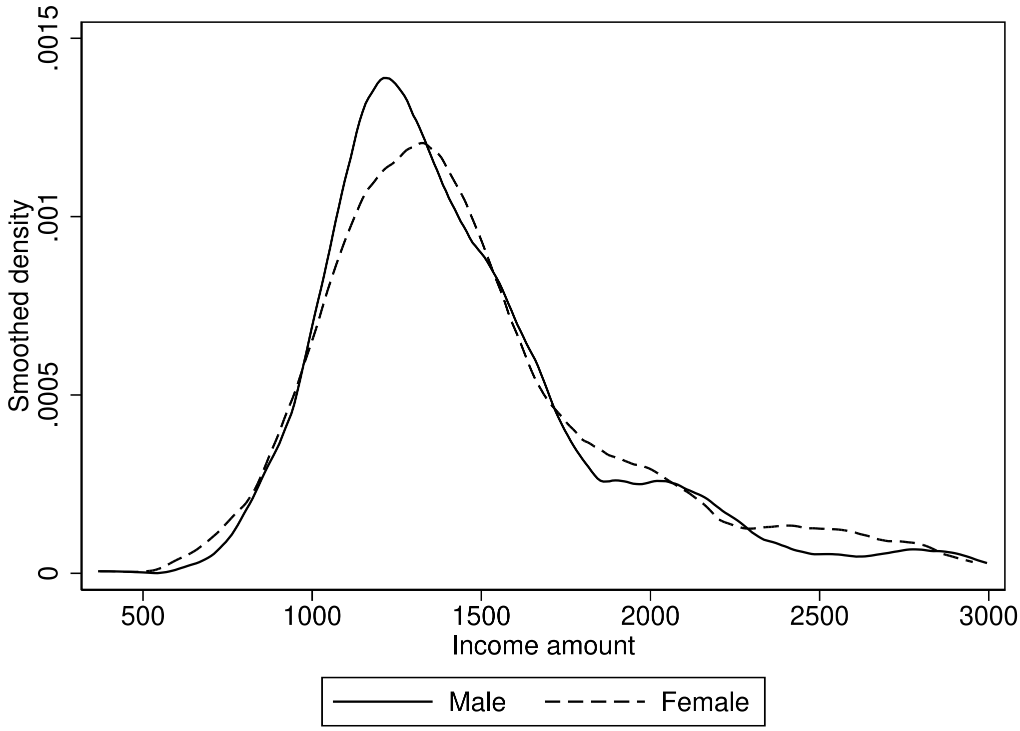
Notes: Lines show the kernel densities of awards (as fixed amounts of money to be transferred by the respondent to the petitioner) set by male and female judges in informal cases.

Figure 1.3: Kernel distribution of calibrated awards in informal cases



Notes: Lines show the kernel densities of the calibrated awards (as a percentage of respondent's income) set by male and female judges in informal cases. A calibrated award is the hypothetical allocation set by judges when the respondent's income is not observable (informal job) based on the judges' revealed allocation preference when the respondent's income is observable (formal jobs). It captures weights assigned by judges to the respondent's number of children inside and outside the trial, and the judge's fixed-effect extracted from formal cases.

Figure 1.4: Kernel distribution of estimated judges' beliefs about the respondent's income



Notes: Lines show the kernel densities of the estimated judges' beliefs about the respondent's unknown income (in local currency) for male and female judges in informal cases. The estimated belief is the ratio of the award (given in absolute terms) divided by the calibrated award (allocation preference - share of the respondent's income to be allocated) when the respondent's income is not observable (informal job).

Figure 1.5: Kernel distribution of petitioner's claims

Notes: Lines show the kernel densities of the petitioner's claim in absolute terms (in local currency) received by male and female judges in informal cases. The claim of petitioners is the amount of money they say the father should transfer to cover the needs of children.

Tables

Table 1.1: Sample characteristics

	All cases (1)	Male judges (2)	Female Judges (3)
Panel A. Hearings			
Settlement (%)	27.4	24.5	29.7
Respondent is formal (%)	20.2	20.8	19.6
Number of judges	149	61	88
Observations	2,371	1,061	1,310
Panel B. Litigations			
Respondent is formal (%)	22.7	22.2	21.3
Number of judges	153	59	94
Observations	1,736	856	880

Notes: This table describes samples corresponding to the two stages (hearing and litigation) of child support cases. Samples contain cases filed in districts with at least one court led by a male judge and one court led by a female judge.

Table 1.2: Balance Table, Case Characteristics by Judge Gender (Hearing Settlement)

	Male Judge		Female Judge		Difference	
	Mean	SE	Mean	SE	Mean	p
Number of children	1.333	0.024	1.326	0.019	0.007	0.816
P. attends (%)	97.8	0.5	98.2	0.4	-0.4	0.591
P.'s attorney attends (%)	67.7	2	64	1.7	3.7	0.168
R. attends (%)	64.6	1.7	65.2	1.6	-0.6	0.785
R.'s attorney attends (%)	35.3	1.7	35.9	1.7	-0.6	0.823
R. is rebel (%)	49.2	2.6	49.8	2.8	- 0.6	0.881
R. is formal (%)	20.8	1.6	19.6	1.4	1.2	0.582
Observations	1,061		1,310		2,371	
Number of judges	61		88		149	

Notes: This table presents a balance table on cases' characteristics. "P." refers to petitioner and "R." to respondent.

Table 1.3: Balance Table, Case Characteristics by Judge Gender (Litigation - Formal)

	Male Judge		Female Judge		Difference	
	Mean	SE	Mean	SE	Mean	p
Number of children (in trial)	1.376	0.048	1.347	0.041	0.029	0.644
Number of children (off trial)	0.318	0.041	0.489	0.087	-0.171	0.079*
Claim (%)	55.5	0.6	56	0.6	-0.5	0.543
P. reports resp.'s income (%)	48	5 3	51.1	4.8	-3.1	0.658
R. reports his income (%)	43.9	4.3	46.6	4.4	-2.7	0.668
P. has assets (%)	0	0	1.1	1.1	-1.1	0.318
R. has assets (%)	4	1.6	5.7	2	-1.7	0.517
Observations	173		176		349	
Number of judges	43		57		100	

Notes: This table presents a balance table on cases' characteristics for cases where the respondent has a formal job. Sample restricted to cases held at districts where there is at least one female and one male judge. "P." refers to petitioner and "R." to respondent.

Table 1.4: Balance Table, Case Characteristics by Judge Gender (Litigation - Informal)

	Male Judge		Female Judge		Difference	
	Mean	SE	Mean	SE	Mean	p
Number of children (in trial)	1.338	0.033	1.329	0.025	0.009	0.810
Number of children (off trial)	0.356	0.045	0.335	0.029	0.021	0.702
Claim (%)	1425.5	97.6	1340.7	57.4	84.8	0.454
P. reports resp.'s income (%)	53.8	3.2	52.8	3.9	1	0.836
R. reports his income (%)	44.3	3.1	45.0	3.1	-0.7	0.866
P. has assets (%)	1.0	0.4	1.2	0.5	-0.2	0.722
R. has assets (%)	6.8	1.3	7.8	0.9	-1.0	0.536
Observations	585		642		1,227	
Number of judges	50		74		124	

Notes: This table presents a balance table on cases' characteristics for cases where the respondent has an informal job. Sample restricted to cases held at districts where there is at least one female and one male judge. "P." refers to petitioner and "R." to respondent.

Table 1.5: Judge's gender effects on child support decisions

	Log(award per child)		Z-score(award per child)	
	Formal (1)	Informal (2)	Formal (3)	Informal (4)
Female judge	-0.068**	-0.059**	-0.25**	-0.16***
	(0.033)	(0.027)	(0.117)	(0.060)
N of children (in trial)	-0.343***	-0.291***	-0.997***	-0.357***
	(0.019)	(0.019)	(0.058)	(0.037)
N of children (off trial)	-0.169***	-0.098***	-0.490***	-0.146***
	(0.011)	(0.016)	(0.043)	(0.028)
Observations	349	1,227	349	1,227
N of judges	100	124	100	124
R2	0.616	0.383	0.540	0.239

Notes: This table presents the estimates of the judge gender-based gap in awards for formal and informal cases. Column 1 uses the log of the award per child in formal cases as a dependent variable. Column 2 uses the standardised award per child in formal cases as a dependent variable. Columns 3 and 4 use the analogue figures for informal cases. Each regression includes district and year fixed effects. Standard errors clustered at the judge level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

Table 1.6: Balance Table, Judges' Characteristics by Judge Gender

	Male Judge		Female Judge		Difference	
	Mean	SE	Mean	SE	Mean	p
Age (years)	43.571	1.105	42.640	0.857	0.931	0.503
Judge is principal (%)	49.2	0.066	35.2	0.051	14.0	0.093*
Years as principal	6.379	1.026	5.871	0.913	0.508	0.712
Wealth (normalised)	0.061	0.118	-0.040	0.114	0.101	0.550
Observations	59		88		146	

Notes: This table shows the balance test for all characteristics of judges available. The value displayed for t-tests are p-values of the difference across groups.

Table 1.7: Pooled OLS estimates with all judges' characteristics

	Z-score(award per child)			
	Formal		Informal	
	(1)	(2)	(3)	(4)
Female judge	-0.247**	-0.261**	-0.157***	-0.161**
	(0.117)	(0.124)	(0.0597)	(0.0685)
Age (years)		-0.00300		-0.0132**
		(0.00774)		(0.00656)
Judge is principal		0.110		0.287***
		(0.114)		(0.106)
Experience as principal (years)		-0.00540		0.00448
		(0.0169)		(0.0125)
Wealth (standardised)		0.0353		-0.114**
		(0.0399)		(0.0537)
Observations	349	334	1227	1172
N of judges	100	98	124	119
R2	0.540	0.554	0.239	0.256

Notes: This table presents the estimates of the judge gender-based gap in amounts for child support when including all judges' characteristics as controls. Columns 1 and 2 use the standardised award per child as a dependent variable in formal cases without and with additional judge's characteristics as covariates, respectively. Columns 3 and 4 use the standardised award per child as a dependent variable in informal cases without and with additional judge's characteristics as covariates, respectively. Each regression controls number of children involved in the trial and other children the defendant has to support. District and year fixed effects are also included. Standard errors clustered at the judge level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

Table 1.8: Marginal effects for settlement

	Likelihood of settlement in hearing session					
	All cases			Both parties attended		
	(1)	(2)	(3)	(4)	(5)	(6)
Female judge	0.0748	0.0777	0.102	0.0755	0.0800	0.100
	(0.0826)	(0.0827)	(0.0862)	(0.116)	(0.116)	(0.120)
N of Children		-0.0624	-0.0528		-0.0520	-0.0359
		(0.0408)	(0.0430)		(0.0493)	(0.0507)
R. is formal			-0.110*			-0.241***
			(0.0637)			(0.0761)
Observations	2,404	2,391	2,253	1,520	1,510	1,419
Pseudo R^2	0.023	0.024	0.027	0.045	0.045	0.052
N. of judges	147	147	142	134	134	129

Notes: This table shows the marginal effects of case characteristics on the likelihood of settlement. Columns 1 to 3 use the whole set of cases. Columns 4 to 6 use the restricted set of cases where both parties attended the hearing session. $Female$ is an indicator variable for whether case was assigned to a female judge. $Children^T$ is a variable containing the number of children involved in the trial. $Formal$ is an indicator variable for whether the defendant has a formal job. Each regression controls for district and year fixed effects. "R." refers to respondent. Standard errors clustered at the judge level in parentheses.* p <0.10, ** p <0.05, *** p <0.01.

Table 1.9: Estimates for settlement agreements

	Amount agreed	
	(1)	(2)
Female judge	-24.95	-24.04
	(27.32)	(28.19)
Respondent is formal	-514.0***	-526.9***
	(17.90)	(19.30)
N of children (in trial)		115.9***
		(19.79)
Observations	653	651
R^2	0.343	0.383
Number of judges	104	104

Notes: This table presents the estimates of the judge gender-based gap in amounts for child support agreed in settlement hearings. Columns 1 and 2 use the log of the amount per child as a dependent variable. Each regression controls for district and year fixed effects. Standard errors clustered at the judge level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

Table 1.10: Gender gap decomposition - pooled OLS estimates

	Total	Preference	Belief
	(1)	(2)	(3)
$Female$	-0.0474**	-0.121***	0.0733**
	(0.0232)	(0.0364)	(0.0364)
$Children^{T}$	-0.295***	-0.352***	0.0571***
	(0.0127)	(0.00817)	(0.0140)
$Children^{-T}$	-0.102***	-0.183***	0.0804***
	(0.0140)	(0.00996)	(0.0113)
Observations	1382	1382	1382
N of judges	107	107	107
R2	0.444	0.820	0.195

Notes: This table presents the estimates of the judge gender-based gap decomposition in informal cases. Columns 1 uses the log of the award per child as a dependent variable. Column 2 uses the calibrated percentage of defendant's income awarded per child as a dependent variable. Columns 3 uses the log of the belief about the defendant's income as a dependent variable. Each regression controls for district and year fixed effects. Standard errors clustered at the judge level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

Table 1.11: Pooled OLS estimates beliefs formation

	Log(belief)	
	Male judge (1)	Female judge (2)
Log of Claim	0.370***	0.253***
	(0.0439)	(0.0341)
Constant	4.692***	5.385***
	(0.292)	(0.218)
Observations	728	641
N of judges	50	59
R2	0.435	0.318

Notes: This table presents the estimates of the belief formation framework. Both columns 1 and 2 use the log of the estimated belief as a dependent variable. Each regression controls for district and year fixed effects. Standard errors clustered at the judge level in parentheses. * p <0.10, ** p <0.05, *** p <0.01.

Chapter 2

Positive Feedback and Score Comparison

I study the effect of providing positive feedback on future participation in mathematical competitions using administrative data on participants in district Mathematical Olympiads in Slovakia. To establish causality, I use a regression discontinuity design in which competitors who achieve a certain score receive a diploma with a "successful" label. I find that positive feedback increases the likelihood of participation in the following category by 3 p.p. (10% percent). Furthermore, by exploiting idiosyncratic variation in score distribution across peer groups, I find that the effect of positive feedback is weaker when surrounded by higher proportions of high-performing peers whereas is not affected when exposed to higher proportions of low-performing peers. This suggests that the value of positive feedback effect is affected by score comparison and this relation is asymmetric.

2.1 Introduction

There is considerable evidence on the important role played by feedback in fostering students' future educational outcomes. For instance, *descriptive* feedback (e.g., absolute test score) increases performance by providing information to students on how their effort translates into performance (Bandiera, Larcinese and Rasul, 2015). Whereas *comparative* feedback (e.g., rank within a group) affects performance and educational attainment not only by providing information (Goulas and Megalokonomou, 2021), but also by activating behavioral forces such as competitive preferences (Azmat and Iriberri, 2010) and social comparison (Dobrescu et al., 2019). However, we know little about the effects of a third type of feedback which adds a level of judgment to students' performance and is ubiquitous in educational settings: *evaluative* feedback or student appraisal (e.g., "you are successful").

In particular, how might evaluative feedback in a competitive setting affect students' willingness to continue competing in the future? Despite the common sense notion indicating that positive (negative) feedback will positively (negatively) affect future participation, the evidence from labour contexts shows that both negative and positive feedback can either increase or decrease motivation. On one hand, individuals react to negative feedback (e.g., "you did badly") by "giving up" or "trying harder" (Cianci, Klein and Seijts, 2010). On the other hand, individuals respond to positive feedback (e.g. "you did well") by "siting on their laurels" or "doubling their efforts" (Dijk and Kluger, 2011). This evidence suggests that the effect might depend on the context and, hence, remains an empirical question. Furthermore, evaluative feedback effects pose an important challenge for research design, since credible identification depends on the ability to isolate exogenous variation in the provision of such feedback while holding constant students' abilities.

This paper adds to this literature by studying how evaluative feedback affects future decisions of children to participate in Mathematical Olympiads. I use administrative records on over 56,000 results from 5th to 9th grade participants at the district Mathematical Olympiads in Slovakia which designs a unique category for each of these grades. I exploit a discrete threshold that determines the provision of positive feedback to students in each category. Specifically, only students who score 9 points or more (in a scale from 0 to 18) receive recognition (a diploma) as "successful" participants. By linking data on participation of individuals in each of these categories over time, I use the feedback rule to construct regression discontinuity (RD) estimates of the effect of receiving positive feedback on the probability of participating in the following category. The advantage of Mathematical Olympiads in quantifying mathematical abilities, the use of cutoffs to determine the provision of feedback, and the inability of either participants or graders to manipulate scores allows for a unique quasi-experiment to test whether evaluative feedback is influencing future participation.

The central finding of this paper is that receiving positive feedback (being labelled as "successful") increases the likelihood of competitors' participation in the following category by 3 percentage points (10% with respect to the baseline). This suggests that positive feedback strongly encourages children to continue training in mathematics. This effect on students' participation might be due not only to responses of students themselves but also of their teachers, and/or parents. Disentangling these reactions is out of the scope of this paper as the current data set does not offer a chance to explore reactions of agents involved in the decision to participate in the Olympiads.

The second contribution of this paper is the study on how score comparisons affect the positive feedback effect on future participation. The transparency of mathematical Olympiad results within districts allows us to investigate whether the feedback effect on subsequent participation differs depending on the absolute performance of their peers at the top and at the bottom of the score distribution. For instance, competitors might lower the value of recognition if their peer with the highest score in the district obtains a much higher score than them or if there are too many high-performing competitors in the district. Conversely, competitors might raise the value of recognition if they learn that their peer with the lowest score in the district obtains a much lower score than them or if there are too many low-performing peers. By exploiting idiosyncratic variation in the tails of score distributions, I find that the effect of positive feedback is weaker in districts with larger highest scores and larger proportions of high-performing peers (e.g., those achieving the maximum score possible) while its effect remains unchanged for districts with less lowest scores and larger proportions of low-performing peers (e.g., those achieving the minimum score possible). These results support the hypothesis that negative comparisons matter more than positive comparisons when valuing positive feedback. This asymmetric relation is consistent with prior research in labour contexts showing that job satisfaction depends on relative pay comparisons, and this relationship is nonlinear (see Hamermesh (2001) and Card et al. (2012)).

Finally, in light of the literature on gender differences in self-confidence and interest in mathematical competitions (e.g., Niederle and Vesterlund (2007) and Croson and Gneezy (2009)), I test whether there are gender differences in the effect of evaluative feedback on future participation in mathematical competitions. I find no evidence that boys and girls react differently to evaluative feedback. Moreover, in this setting, I find that girls enter into competition as much as boys do. These findings are at odds with the literature on gender differences in competitive traits, usually based on adolescents and adults, and suggest that gender differences are not relevant either for children or the selected sample of highly accomplished students.

This paper contributes to two strands in the literature of Economics of Education. First, the central finding of this paper builds on an extensive literature studying the effects of feedback on educational outcomes. A comprehensive review is available in Villeval (2020) who shows that the literature has focused on studying descriptive and comparative feedback. This paper provides evidence on evaluative feedback which has a less objective nature in comparison to descriptive and comparative feedback, and has been less explored in educational settings. To the best of my knowledge, only two papers have partially addressed the effects of evaluative feedback in educational contexts. First, Bedard, Dodd and Lundberg (2021) evaluate whether positive feedback on performance in first economics courses at university increases the probability of majoring in Economics. The limitation of this field experiment is that the difference between the feedback given to the treatment group ("strong performance in Economics 1") and to the control group ("successfully completing Economics 1") is inconspicuous in comparison to the setting of this study (positive feedback versus no feedback). Second, Hoogveld and Zubanov (2017) study the effect of recognition on performance involving also first-year university students and find that the recipients of recognition did not do better, while the non-recipients significantly improved their performance. Although designed as a field experiment, their main results are based on an RD approach with a restricted sample size, which poses concerns on their external validity. In contrast, this paper provides credible field-based evidence of the importance of evaluative feedback for educational outcomes and its results are robust to all sorts of different specifications.

This study also contributes to recent literature that studies the effect of failure in competitions on future participation. These studies conduct RD estimations relying on score cutoffs that determine failure at advancing to further rounds. Buser and Yuan (2019) use data on Dutch mathematical Olympiads to estimate the effect of losing relative to winning (round 1 of national Olympiads) on subsequent participation in the competition. Ellison and Swanson (2021) address the same research question in the context of the American Mathematical Competitions (AMC). Both studies find evidence that competitors react negatively to losing relative to winning. However, despite the appeal of an RD design, the interpretation of the estimates of these studies depends critically on the treatment given to students who narrowly pass the score cutoff. In particular, in both settings, there is a multi-stage structure such that students who reach a certain score in a given round will proceed to a next round for which they will train more, gain more competition experience, and even additional recognition if passing a certain score in later rounds. This raises the question: to what extent is the effect driven by failure/feedback? In my setting, evaluative feedback is isolated from additional experiences since Olympiads for early grades have no multi-stage structure and, therefore, results can be attributed to it.

As results suggest that children infer their capacity to do mathematics not only from their absolute and relative performance but also from evaluative feedback, this study poses the following question for policy matters: shall we design evaluative feedback structure that maximizes interest and participation in mathematics for all students, or one that focuses on talented students and relegates the less skilled at early ages? It seems that the design of labeling participants at early ages might be unnecessary as there are no further rounds in the competition. Moreover, such institutional design might have massive consequences on children's encouragement since their non-cognitive traits such as grit and conscientiousness, relevant for this matter, are less developed than those of adolescents and adults (Mike et al., 2015) and, therefore, are more prone to be affected by evaluative feedback. However, prescriptions in this regard deserve further lines of research and normative analysis. The institutional design of feedback for talented young students has more complex consequences as there is evidence that the development of their mathematical skills expands the knowledge frontier later in life (Agarwal and Gaule, 2020*a*).

The paper is structured as follows. Section 2.2 describes the Mathematical Olympiads and data used. Section 2.3 presents the econometric method, main results and heterogeneous effect analysis by gender of participants. Section 2.4 addresses score comparisons. Section 2.5 concludes.

2.2 Institutional Background and Data

2.2.1 Background

Mathematical Olympiads are competitions usually held on the basis of regional and national rounds within countries whose ultimate goal is to select the best students to represent them at the International Mathematics Olympiad (IMO). Olympiads are proof-based contests consisting of few problems (at the IMO, 6 questions to be solved in two days) drawn from geometry, number theory, algebra, and combinatorics. There is strong evidence the mathematical Olympiads are reliably capturing math abilities of students (Ellison and Swanson, 2010*a*) and that IMO scores are highly predictive of math publications and citations twenty years in the future (Agarwal and Gaule, 2020*a*).

Although these competitions are aimed at high school students for the reasons mentioned, Mathematical Olympiads in Slovakia target students at primary school levels as well. This special feature in Slovakia makes competitions at this level suitable for studying pure feedback effects. Indeed, from the 5$^{\text{th}}$ to 8$^{\text{th}}$ grades, students only compete in a district round, as opposed to students from the 9$^{\text{th}}$ to 13$^{\text{th}}$ grades, where additional rounds (regional and national) are added. As already discussed in the introduction, a multi-stage structure complicates the analysis of the effect of scoring above the cutoff as it implies that competitors not only receive evaluative feedback but also gain additional training for competition in later rounds. Therefore, this analysis focuses on competitions designed for 5$^{\text{th}}$ to 8$^{\text{th}}$ grade students. Olympiads for elementary grades are categorised by adding the prefix Z to the corresponding grade. For instance, the Olympiad category for 5$^{\text{th}}$ grade students is called "Z5".

Contests for primary levels are organised by the Slovak Committee of Mathematical Olympiad (SKMO) every year at the district level. To do so, district level committees are formed to manage the competitions locally. Tests for 5$^{\text{th}}$ to 8$^{\text{th}}$ grade students consist of 3 questions (6 points each), while the 9$^{\text{th}}$ grade test involves 4 questions (6 points each). Crucially, score thresholds are established by the SKMO in each category. For Z5, Z6, Z7 and Z8, students who score at least 9 points are recognised as "successful". According to the SKMO, recognition diplomas are given to these students, while no recognition is given to students who score 8 points or less[1]. For Z9, the same recognition is given at 12 points. These thresholds have been constant during the period 2011-2018. Finally, it is important to note that district committees provide descriptive (total score), comparative (rank within the district) and evaluative feedback ("successful") to all participants within the district. This implies that a given participant is not only informed about his/her own score/rank/feedback but also about the score/rank/feedback of each participant in the district.

[1]SKMO informed that the feedback provided to students who score 8 points or less depend on the tutor. As a general recommendation, teachers do not say to students that they were "unsuccessful". Instead, some teachers might deliver verbal messages encouraging these competitors to participate the next year.

2.2.2 Data and Analysis Samples

I collected data on participants at the district mathematical Olympiads between 2011 and 2017 in categories Z5 to Z8. The panel structure allows us to track those competitors and observe whether they participate in the next corresponding categories (Z6 to Z9) during 2012-2018. To have a longitudinal structure, I built unique identifiers for students and schools based on names provided and consistency over time. This data includes grade (5^{th} to 9^{th}), gender[2], language (Slovak or Hungarian)[3] and type of school (regular or grammar school)[4]. For each participant in categories Z5 to Z8, I also observe the score obtained and the rank within the district for the correspondent category. For example, I can track competitor i in category c with score s_i and rank r_i at district d in year t, and observe whether he/she participates in the following category in year $t+1$.

Table 2.1 shows characteristics of all participants (column 1), as well as characteristics of participants in the RD analysis samples (columns 2–3). Panel A shows characteristics for all participants in any category. As shown in column 1, for all 46,968 tests considering all categories (Z5 to Z8), 37 percent of students participated already in a previous category on average, almost half of participants are girls, 8 percent of students are taught in Hungarian language, and 14 percent of students are enrolled in a grammar school. Panels B, C, D, and E correspond to samples for categories Z5, Z6, Z7, and Z8, respectively. Two important differences across categories are worth mentioning. First, it can be seen that the higher the category, the lower the amount of total observations. For instance, I observe 16,334 participants in the Z5 category, and 8,566 participants at Z8 category. Second, the higher the category, the higher the proportion of students with previous experience. For instance, in Z5, no student has previous experience in Olympiads as they only start in 5^{th} grade. In Z8 category, 66% of participants attended the Olympiads in previous categories at least once. However, it is interesting to note that most of the characteristics remain the same in each category. Interestingly, the proportion of girls participating represents around half of total participants in any category.

I construct two RD samples for analysis. Based on the cutoff score that determines the type of feedback, I select students who score within 2 points

[2] I inferred the sex of participants based on their names.

[3] Students who are taught in Hungarian language are identified by their school description.

[4] Like the language classification, students enrolled in regular or grammar schools are identified by their school description.

(column 2) and 3 points (column 3) of the cutoff. For instance, for column 2 we have 11,182 competitors who scored from 7 to 10 points. It can be seen that for any of these two windows, students do not vary along characteristics (past participation, proportion of girls, students taught in Hungarian language, and students enrolled in grammar schools). Moreover, students of the RD samples are comparable to students of the whole sample along these characteristics. Given the discrete nature of the running variable, an RD sample based on competitors scoring 8 or 9 points does not allow us to conduct the analysis since the treatment and points are perfectly correlated. In other words, it is not feasible to attribute changes in the outcome to the treatment (positive feedback).

2.3 RD-based Analysis of Positive Feedback

2.3.1 Research Design

To evaluate the effects of receiving positive feedback on subsequent participation at the mathematical Olympiads, I adopt a sharp regression discontinuity (RD) design around the score cutoff of 9 points that determines the provision of positive feedback. In particular, we estimate the following model:

$$Y_{i,c+1} = \alpha_1 \mathbf{1}(S_{ic} \geq 9) + f(S_{ic}) + \mathbf{1}(S_{ic} \geq 9) \times f(S_{ic}) + X_i'\gamma + \theta_{dtc} + \theta_s + u_{i,c+1}, \quad (2.1)$$

where $Y_{i,c+1}$ is an indicator variable for participating the next category $c+1$. $\mathbf{1}(S_{ic} \geq 9)$ is an indicator variable for scoring equal to or above the threshold in category c. $f(S)$ is a polynomial function of the number of points scored S. X_i is a vector of controls for gender, past participation, rank within the district in category c. θ_{dtc} is a district-by-year-category fixed effect that controls for any unobserved shocks common to all competitors within a district, year and category. Finally θ_s is a school fixed effect.

The parameter of interest α_1 provides an estimate of the causal effect of receiving positive feedback on future participation averaged across all categories. Under the identification assumption that u_i does not change discontinuously at 9 points, this regression identifies an unbiased estimate. Intuitively, the assumption is that the score at a given category is smoothly related to characteristics that affect participation in the following category, which implies that $f(S)$ is constant in a neighbourhood around the 9 points threshold. Therefore, competitors who scored just below 9 points are a suitable control group for individuals with scores just above 9 points, and any difference in their participation in the following category can be attributed to the fact that they receive positive feedback. Formally, $f(S)$ is nonparametrically identified at $S = 9$ (Hahn, Todd and der Klaauw, 2001).

To estimate the model in equation (2.1), I use three different bandwidths of 1, 2, and 3 points to the left and the right of the threshold, and I adopt two different approaches. First, I assume that shocks in districts, years, and categories work independently and include separate fixed effects for each level (θ_d, θ_t and θ_c). Second, I relax that assumption and allow for shocks to be specific for a given district at a particular year in a specific category (θ_{dtc}). Although this model rules out all confounders at the district-year-category level, is more demanding because 1,787 fixed effects need to be estimated, compared to 78 district, 8 year, and 4 category fixed effects in a model with separate fixed effects.

2.3.2 Validity of RD Design

In our context, the ability of competitors or scorers to manipulate which side of the score threshold would fall might be a concern. In this regard, it must be assumed that it is random for a competitor to have a score either just below or just above the score cutoff in order to establish identification.

There are a number of reasons that this assumption is likely reasonable. The first is the level of accuracy of mathematical Olympiads to measure mathematical abilities (Agarwal and Gaule, 2020*a*). This implies that students with small score differences are virtually equally talented. Second, students do not know the scoring criteria in advance, meaning that it is implausible for them to manipulate their answers and choose whether to get negative or positive feedback. In fact, it is fair to assume that all students are aiming to score the highest points. Third, scorers follow a national grading rule and do not have incentives to manipulate the score of students who are just below the threshold that determines the type of feedback received. Indeed, district rounds are not followed by other rounds and, therefore, grading participants below or above the cutoff do not involve costs for the organisation. Moreover, scorers do not tutor participants so there is no incentive to encourage particular participants via score manipulation.

Figure 2.1 contains a histogram displaying the number of observations in each possible score at the mathematical Olympiads. It includes results from all categories (Z5 to Z8) during 2011–2017. The distribution of scores shows no evidence of endogenous sorting to one side of either of the threshold studied. Figure 2.2 shows the same analysis by category. Likewise, there are no concerns of score manipulation for any category.

It is interesting to observe heaps at 0, 6, 12 and 18 points in Figure 2.1. The fact that a good proportion of competitors scored 0 points might indicate that they were pushed by their parents without preparing for the examination. However, the other heaps might be explained by the fact that some competitors only aim at answering a question for which they know the complete proof and are not interested in providing wrong or incomplete proofs. This raises another type of concern by which competitors around the threshold might have different strategies to solve the test: some competitors might provide only complete proofs while others might be willing to provide incomplete proofs in order to obtain a higher score.

Finally, I conduct an additional test of sorting by examining regression models based on equation (2.1) with the students' characteristics used as the dependent variables which should remain unchanged at the thresholds. The participants' demographics I examine are gender and language of instruction (Slovak or Hungarian). Table 2.2 shows estimates of the effect of receiving positive feedback on these predetermined characteristics. The regression models estimated employ bandwidths of 2 and 3 points away of the cutoff (9 points). For each student's demographics, all fixed effects (district, year, category and school) are included in the regressions.

Table 2.2 shows that overall the predetermined individuals' characteristics (gender and past participation in the Olympiad) are not statistically related to receiving positive feedback. With the exemption of column (2), these characteristics are balanced between control and treatment groups. Further graphic evidence for the previous analysis is presented in Figure 2.3, which shows the relationship between students' score and predetermined students' characteristics. It shows bins of predetermined characteristics and corresponding fitted regression lines based on equation (2.1) which should remain unchanged across the scoring thresholds. Figure 2.3 shows that demographic factors are stable across the feedback threshold. The stability of predetermined characteristics gives additional credibility that the regression discontinuity design can deliver unbiased estimates in this context.

2.3.3 Impact on Future Participation

In this subsection I examine the effect of receiving positive feedback on future participation in the mathematical Olympiads. An advantage of the RD design is that it provides a graphical depiction showing how the positive feedback effect is identified. Thus, I begin with a graphical analysis of the positive feedback effect before turning to a more detailed regression-based analysis.

Figure 2.4 plots the fraction of participants in year t who participate in year $t+1$ by score obtained in t, and predicted participation rates based on simple regression models for all participants on each side of the cutoff score. Each observation is the proportion of competitors participating the year after in score bins. The running variable, score obtained in year t, has been normalized so the cutoff (nine points) is displayed at zero. Thus, the black lines represent the fitted regressions in the intervals -9 to -1, and 0 to 9.

Before focusing on the discontinuity, it is worth noting the strong relationship between the score obtained in a given year and the likelihood of participating in the following category as shown in Figure 2.4. The higher the score obtained, the higher the likelihood of subsequent participation. For instance, around 20 percent of students who scored 0 points (or -9 points below the cutoff) participate in the following category, while around 70 percent of students who scored 18 points (or 9 points above the cutoff) do so. Moreover, it can also be seen that the higher the points, the larger confidence intervals of the mean of future participation. This is explained by the fact that students scoring high points are fewer as shown in Figure 2.1.

Figure 2.4 reveals a sharp jump in the fraction of students participating in the following category at the cutoff, with competition participation rising from 41 percent to 51 percent. This graph provides strong evidence that positive feedback has large effects on students' willingness to compete again. While the regression lines illustrate this relationship at the cutoff score, the unrestricted fraction means indicate the underlying noise in the data. As can be seen, on each side of the cutoff score, the relationship between score at t and participation at $t+1$ is smooth providing strong evidence that no other factor is affecting participation at the cutoff apart from feedback. As a reminder, these students are all labelled as "successful" depending on which side they are from the cutoff.

Figure 2.5 contains the same plot for each of the four transitions (from Z5 to Z6, from Z6 to Z7, from Z7 to Z8, and from Z8 to Z9) separately. It can be seen that in all transitions there is a discontinuous jump in the probability of participating the next year when competitors fall just short of the cutoff, except for in the Z6 category. However, Figures 2.4 and 2.5 should be taken carefully as they show the unconditional subsequent participation and do not include any control. In particular, they do not take into account that some students have participated before the period of analysis which might influence future participation for students around the threshold. For instance, in Panel B (transition Z6 to Z7), as shown in Table 2.1, 46 percent of these competitors participated in the Z5 category. Only Panel A of Figure 2.5 does not suffer from this issue: for the transition from Z5 to Z6, no student has previous experience in the Olympiads as they do not exist for lower grades.

Having shown the raw patterns of future participation around the score cutoff, I present regression-based estimates. Table 2.3 shows the estimated effect of positive feedback on future participation. I present estimates for two bandwidths as described in Table 1: +/-2 points; +/-3 points away from the cutoff. The discontinuity estimates are between 3.4 p.p. or around 10% with respect to the baseline (column 2) and 5.3 p.p. or around 15% (column 1). This means that competitors react positively to positive feedback. Table 2.3 also shows the estimated effect based on regressions with separate fixed effects (columns 1 and 3) and district-by-year-by-category fixed effects (columns 2 and 4). It can be seen that the estimated effect in the latter is slightly lower and consistent in all RD samples. Because of its consistency and virtue to account for all non-observable factors that might affect future participation at the district-year-category level, this model is my preferred specification.

This study has two important limitations. First, the data set does not allow us to disentangle the reactions of students from the responses of their tutors and parents. Indeed, the decision to participate is voluntary and might be based on interaction between students, tutors and parents. For instance, tutors might be more encouraging, offer better training and/or set higher expectations for students who receive positive feedback, while parents might reward them and/or push their children to continue training. Understanding these mechanisms is important in designing feedback policies. Second, I cannot identify how tutors manage the feedback given to participants who achieved a score just below the threshold as this information is not available. In this regard, tutors might either provide verbal encouragement or negative feedback to these students. With respect to the latter, some studies suggest that negative feedback may increase motivation and future performance (Cianci, Klein and Seijts, 2010) and, therefore, one could expect that the estimated effect in this study may be a lower bound.

2.3.4 Gender Differences in Reaction to Feedback

In this subsection, I investigate whether there are gender differences in the reaction to feedback. Thus, the main interest is not the discontinuity itself (the effect of receiving positive feedback on future participation), but whether there are gender differences in the discontinuity. In other words, whether boys and girls react differently to positive feedback. Therefore, I estimate the following equation:

$$Y_i = \alpha_1 T_i + f(S_i) + f(S_i) \times T_i + \alpha_2 F_i + \alpha_3 T_i \times F + f(S_i) \times F + f(S_i) \times F \times T_i + u_i, \tag{2.2}$$

where Y_i is an indicator variable for participating the next year, T_i is analogous to the indicator variable $\mathbf{1}(S_i \geq 9)$ in equation (2.1) which indicates whether participant i scored equal or above the threshold, $f(S)$ is a polynomial function of the number of points scored S. The parameter of interest in this analysis is α_3, which estimates the gender difference in reaction to feedback. Following Buser and Yuan (2019), I include two important interactions allowing for different slopes for each gender: the polynomial $f(S)$ with the participant's gender F; and the triple interaction between the polynomial $f(S)$, the treatment T, and the participant's gender F. Again, I run this analysis using separate fixed effects and district-by-year-by-category fixed effects.

Like the general analysis, I first provide graphical evidence of gender differences of the impact of positive feedback on future participation in mathematical Olympiads. Figure 2.6 plots the fraction of participants in year t who participates in year $t+1$ by score obtained in t for boys and girls. First of all, it is remarkable that conditional on the score obtained in a given year, there is no evidence of gender differences in subsequent participation. To the left (less skilled) and the right (more skilled) of the cutoff, both boys and girls participate in Olympiads the year after. Regarding the gender difference around the cutoff, Figure 2.6 shows no graphical evidence of different reactions to positive feedback.

Next, I test gender differences in reaction to positive feedback by estimating equation (2.2). 5^{th} row in Table 2.4 shows the estimates of the difference between boys and girls in the effect of positive feedback on future participation. It presents the estimates for two different ranges of score around the cutoff: 2

points (columns 1 and 2) and 3 points (columns 3 and 4). The sixth row in Table 4 reveals that the difference is not statistically significant except for one bandwidth (+/- 2 points) at the 10% level when allowing for district-by-year-by-category fixed effects. The results taken together suggest that there are no gender differences in reaction to positive feedback.

The evidence of gender differences from this study is striking. First of all, the fact that the sample is gender balanced goes in contrast with the consensus on the lower willingness of girls to compete in mathematical subjects (Niederle and Vesterlund, 2010). As shown in Table 2.1, 47 percent of all tests were taken by girls. Second, it shows that there is no evidence that boys and girls react differently to feedback, which also goes in contrast against similar studies (see Buser and Yuan (2019) and Ellison and Swanson (2021)). As previously explained in the introduction, although the setting of this study (one stage structure) is not comparable to theirs (multi-stage structure), this study provides evidence that such differences are not important in the context of high achievers and/or children.

2.4 Positive Feedback and Score Comparison

In this section I investigate whether the positive feedback effect is affected by score comparison. As explained in section 2.2.1, the SKMO disseminates results such that everyone is able to observe the score of everyone else in a district (see Appendix A.3 and Appendix A.4 for examples). I investigate whether this transparency can affect the feedback effect on subsequent participation through score comparison. Individuals who scored just above the cutoff (9 points) might feel differently about the recognition received (positive feedback) depending on being exposed to high- and low-performing peers in their districts. For instance, if they learn that many peers achieved high scores, they might assign lower value to positive feedback. In contrast, if they learn that several peers obtained low scores, they might give more value to positive feedback.

Section 2.4.2 provides a simple graphical analysis of the effect of positive feedback on subsequent participation in districts with different maximum and minimum scores achieved. Since characteristics of the score distribution in a given district might be correlated with the probability of receiving feedback, I propose two causal approaches in section 2.4.2 to inspect how high- and low-performing peers can affect the value of positive feedback on encouraging competitors by exploiting idiosyncratic variation in score distributions.

2.4.1 Descriptive Evidence

To illustrate, consider two districts “A” and “B” with different score ranges. In district “A”, the maximum score achieved is 12 points, while in district “B”, is 18 points. The question is whether the effect of receiving positive feedback is different in districts “A” and “B”. The hypothesis is that the effect of positive feedback is less strong in district “B” than in “A”, as students who barely passed the cutoff might feel that the recognition (positive feedback) is not too valuable, as there is one competitor who did extremely well in comparison to them. Analogously, if in district “A” the minimum score achieved is 0 points, while in district “B” it is 7 points, one can expect the effect of positive feedback to be less strong in district “B” than in “A” as competitors who barely passed the cutoff might feel that the recognition is not too valuable as there is least talented competitor almost obtained the same recognition.

I start by providing graphical evidence by grouping districts with respect to the maximum and minimum scores achieved and show the effect of evaluative feedback separately. To simplify the analysis, I set different breaking points for maximum scores (13, 14, and 15 points) and minimum scores (3, 4 and 5) to classify districts. For the maximum score analysis (13 points breaking point), I compare the positive feedback effect among competitors from districts where the maximum score achieved was 12 or less, versus its effect among competitors living in districts where the maximum score achieved was 13 or more. Likewise, for the minimum score analysis (3 points breaking point), I compare the positive feedback effect among competitors in districts where the minimum score achieved was 2 or less, versus its effect among competitors from districts where the minimum score achieved was 3 or more. According to our reasoning, for both analyses, the positive feedback should have a stronger effect among competitors from the first group of districts than from the second group. Table 2.5 and 2.6 describe the samples used for the analysis.

Panels (a) and (b) of Figure 2.7 show the effect of positive feedback in districts where the maximum score was 12 or less (equivalently, distance from the cutoff being 3 or less) versus its effect in districts where the maximum score was 13 or more (equivalently, distance from the cutoff being 4 or more), respectively. Naturally, in panel (a), the fitted line to the right of the cutoff is based on only 4 points, while in panel (b) there is no such restriction. The same applies in panel pairs (c) and (d), and (e) and (f) for 13 and 14 points as breaking points to group districts, respectively. Figure 2.7 suggests that the effect of positive feedback in districts where there are relatively less talented competitors is around twice that in districts where there are relatively more talented competitors. This suggests that positive feedback is affected negatively when recipients are surrounded by high-achievers.

Panels (a) and (b) of Figure 2.8 show the effect of positive feedback in districts where the minimum score was 2 or less (equivalently, distance from the cutoff being -7 or less) versus its effect in districts where the minimum score was 3 or more (equivalently, distance from the cutoff being -6 or more), respectively. By construction, in panel (b), the fitted line to the left of the cutoff is based on only 6 points, while in panel (a) there is no such restriction. The same applies in panel pairs (c) and (d), and (e) and (f) for 4 and 5 points as breaking points to classify districts, respectively. Figure 2.8 suggests that there is no difference between the effect of positive feedback in both groups. As described in Table 6, there are few districts where the minimum score achieved was, for instance, 5 points or more (comprising only 1,085) which explains the noise shown in panel (f). Although Figure 2.8 suggests that positive feedback is not affected when surrounded by low-achievers, I proceed to test it using a regression model.

The limitation of this approach is that the score composition within a district can itself affect a competitor's performance and their likelihood to receive positive feedback. In other words, competitors from districts with different score distributions are not comparable. For instance, Table 2.5 shows that 46% of competitors in districts where the maximum score achieved is 12 points have some previous experience at the Olympiads while the analogous figure for competitors in districts where the maximum score achieve is 13 points or more, is 35%. The same issue applies for the division of competitors based on different minimum score achieved in their districts.

More generally, the challenge of identifying the effect of being exposed to high- and low-performing peers on the value of positive feedback relates to the fact that competitors are self-selected to live in districts based on their quality of schools and, therefore, the value of being recognised in mathematical Olympiad is mechanically related to the characteristics of their peers. To overcome this issue, I develop two different but related causal approaches to hold peer abilities constant and focus on "as good as random" differences in low and high points of reference.

2.4.2 Exposure to Minimum and Maximum Scores

The first approach consists of capturing "as good as random" differences in minimum and maximum levels of scores achieved across districts. To account for observed and unobserved characteristics of districts and competitors that might be correlated with minimum and maximum scores achieved, I exploit idiosyncratic variation in the tails of score distributions across district-year and category levels. By doing so, I compare the effect of receiving positive feedback on competitors that face the same environment, except for the fact that in certain districts they are exposed to a higher/lower maximum and minimum scores achieved as a result of idiosyncratic variation.

The key identifying assumption in this approach is that changes in the maximum and minimum scores achieved in a district-year are uncorrelated with observed and unobserved factors that could themselves affect the likelihood of receiving positive feedback. To test for this, I check whether positive feedback given to competitors around the threshold explains different levels of maximum and minimum scores achieved in their districts. Table 2.7 shows the results of conducting this analysis as in equation (1). For instance, in columns (1) and (2), the maximum and the minimum scores achieved in the district are the dependent variables using the +/- 2 points bandwidth, respectively. Table 2.7 shows that there is no evidence that the competitors above the threshold are more or less exposed to different levels of maximum and minimum scores than competitors below the threshold.

I study additional gains or losses on the positive feedback effect for individuals in districts with different minimum and maximum scores achieved as follows:

$$\begin{aligned} Y_{id,c+1} &= \alpha_1 T_{idc} + f(S_{idc}) + T_{ic} \times f(S_{idc}) \\ &\quad + \beta_1 S_{dc}^{max} + \beta_2 S_{dc}^{min} + \beta_3 T_{idc} \times S_{dc}^{max} + \beta_4 T_{idc} \times S_{dc}^{min} + X_i' \gamma + u_{id,c+1}, \end{aligned} \tag{2.3}$$

where $Y_{id,c+1}$ is an indicator variable for individual i in district d participating in the next category $c+1$. $T_{idc} = \mathbf{1}(S_{idc} \geq 9)$ is an indicator variable for scoring equal or above the threshold in category c. S_{idc} is the number of points scored. S_{dc}^{max} is the maximum score achieved in district d at category c. S_{dc}^{min} is the minimum score achieved in district d at category c. Finally, X_i is a vector of controls for gender, past participation, rank within the district in category c. Finally, all regressions are estimated under two models: using separate fixed effects and district-by-year and category fixed effects. For this analysis, β_3 and β_4 are the coefficients of interest as they capture the gain/loss of the positive feedback effect on subsequent participation when surrounded by better peers below and above the cutoff, respectively.

Table 2.8 shows the results of estimating equation (3) using the two RD samples used throughout this paper. For instance, column (1) shows the estimations for the sample using a bandwidth of $+/-2$ points and separate district, year and category fixed effects. The interpretation for the interaction between treatment and the maximum score achieved in the district is as follows: an increase of 1 point in the maximum score achieved in the district is associated with a reduction of 1.1 percentage points of the positive feedback effect on subsequent participation (row 6). However, I find that an increase in one point in the minimum score achieved in the district does not affect the effect of positive feedback (row 7). Rows 6 and 7 in columns (3) and (4) show the same results for the +/-3 points bandwidth. This evidence suggests that positive feedback is associated with a lower effect on subsequent participation when surrounded by better peers above the cutoff and no gain/loss when surrounded by better peers below the cutoff.

2.4.3 Exposure to High- and Low-performing Peers

The second approach is similar to Mouganie and Wang (2020) and accounts for observed and unobserved characteristics of districts and competitors that might be correlated with high- and low-performing peer composition by exploiting idiosyncratic variation in the distribution of test scores across districts and cohorts. Thus, I compare the effect of receiving positive feedback on competitors that face the same environment, except for the fact that in certain peer groups they are exposed to a higher proportion of high-performing or low-performing peers as a result of idiosyncratic variation.

I conduct this analysis using three different definitions of high- and low-performing peers. In the most stringent definition, a high-performing peer is the competitor who scored 18 points (the maximum possible score) while a low-performing peer is defined as the peer who scored 0 points (the minimum possible score). The alternative pair of definitions for high- and low-performing peers correspond to 17 and 1 points, and 16 and 2 points, respectively.

Similarly, I first check whether positive feedback given to competitors around the threshold is related to different proportions of high- and low-performing peers. Tables 2.9, 2.10 and 2.11 show the results of conducting this analysis as in equation (1) for all the three definitions of high- and low-performing peers. For instance, in columns (1) and (2) of Table 2.9, the proportion of high-performing peers (those who obtained 18 points) and the proportion of low-performing peers (those who obtained 0 points) in the district are the dependent variables using the +/- 2 points bandwidth, respectively. Tables 2.9, 2.10, and 2.11 show that there is no evidence that the competitors above the threshold are exposed to different proportions of high- and low-performing peers than competitors below the threshold.

Thus, by exploiting random exposure to high- and low-performing peers, I estimate the following equation:

$$\begin{aligned} Y_{id,c+1} = {} & \alpha_1 T_{idc} + f(S_{idc}) + T_{ic} \times f(S_{idc}) \\ & + \beta_1 H_{dc} + \beta_2 L_{dc} + \beta_3 T_{idc} \times H_{dc} + \beta_4 T_{idc} \times L_{dc} + X_i' \gamma + u_{id,c+1}, \end{aligned} \tag{2.4}$$

where $Y_{id,c+1}$ is an indicator variable for individual i in district d participating in the next category $c+1$. $T_{idc} = \mathbf{1}(S_{idc} \geq 9)$ is an indicator variable for scoring equal or above the threshold in category c. S_{idc} is the number of points scored. H_{dc} and L_{dc} are the proportion of high-performing and low-performing peers in category c at district c, respectively. Finally, X_i is a vector of controls for gender, past participation, rank within the district in category c. Similar to the previous strategy, I include fixed effects at the district-year level to account for factors that would impact all competitors in a district-year.

Table 2.12 shows the results on how the effect of positive feedback is affected by being exposed to a higher proportion of high- and low-performing peers. Columns (1) and (2) show estimates from the most stringent definition of peers' type of performance using 2 and 3 points as bandwidths, respectively. These estimates indicate that exposure to more high-performing peers decreases the positive effect of feedback while exposure to low-performing peers has no effect. In column (2), the estimate of -0.027 in the 6$^{\text{th}}$ row indicates that a 1 standard deviation increase in the share of high-performing peers decreases the effect of positive feedback on subsequent participation by 2.7 percentage points. In contrast, the estimate of -0.008 in the 7$^{\text{th}}$ indicates that the effect of positive feedback is unaffected by being exposed to a larger number of low-performing peers. The estimate for the interaction between receiving positive feedback and the exposure to high-performing peers is statistically significant at the 5% level for the 2 points bandwidth, and at the 1% level for the 3 points bandwidth. These results are robust to all definitions of high- and low-performing competitors, and the inclusion of separate district and year fixed effects (see Table 2.13).

These results support the hypothesis that negative comparisons (with respect to larger maximum scores or higher proportion of high-performing peers) matter more than positive comparisons (with respect to lower minimum scores or higher proportion of low-performing peers) for a competitor's encouragement. This asymmetric relation is consistent with prior research in labour contexts showing that job satisfaction depends on relative pay comparisons, and this relationship is nonlinear (see Hamermesh (2001) and Card et al. (2012)).

2.5 Conclusions

I study the effect of evaluative feedback on future participation in mathematical Olympiads. By using a regression discontinuity design based on a strict cutoff that determines the provision of feedback, I credibly isolate the effect of positive feedback from other types of feedback (descriptive and comparative), and observable and non-observable factors influencing future participation. I find that labelling children as "successful" positively affects their subsequent participation in contests.

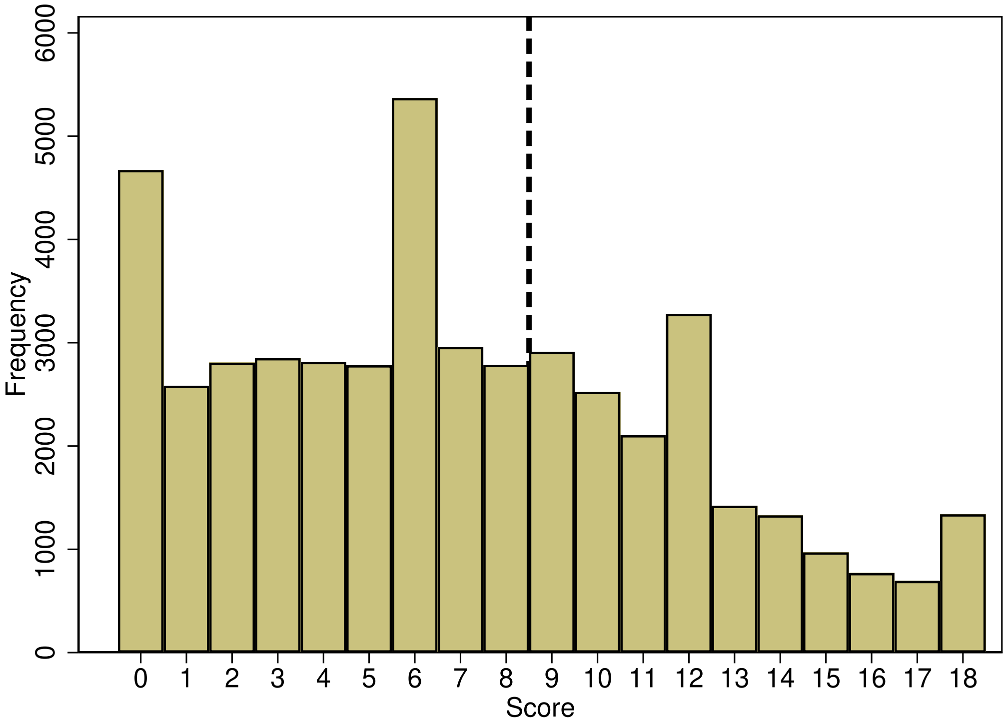
Furthermore, I provide evidence that the magnitude of the effect of evaluative feedback is affected by score comparison. By exploiting idiosyncratic variation in the tails of score distributions, I find that the effect of positive feedback is lower in districts with a larger maximum score and higher proportion of high-performing peers while it does not experience gain or loss in districts with a lower minimum score or higher proportion of low-performing peers. These results suggest that the competitors weight positive feedback by comparing themselves with their peers who are at the top but not with the ones at the bottom of the score distribution.

Finally, I find no gender differences in the effect of evaluative feedback on future competition, meaning that both boys and girls are equally affected by this type of feedback. Moreover, descriptive evidence shows that boys and girls almost equally participate in mathematical competitions, and they dropout at the same rates conditional on past performance. These results are in contrast with the vast literature pointing out gender differences in competitive traits. I speculate that one potential explanation for this divergence lies on the fact that this study is based on children whose traits are not yet affected by gender norms, while most of this literature is based on studies with adolescents and adults.

Although this type of feedback is ubiquitous in many educational settings, it has been relatively unexplored in the literature of Economics of Education. The results of this study show that evaluative feedback has a large effect on engaging children's interest in mathematics and indicate this type of feedback deserves further research. Moreover, this study also opens new research avenues about the design of feedback given in schools.

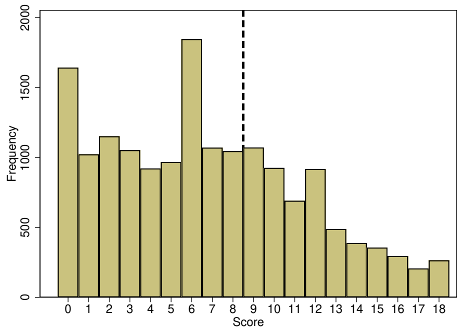
Figures

Figure 2.1: Histogram of Running Variable for RD Analysis

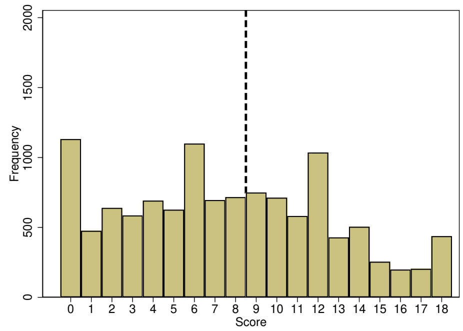


Notes: Sample is all student-test results from categories Z5 to Z8 between 2011-2017 (as in Table 2.1, column 1). Bars reported within bins of width 1. Sample size is 46,968.

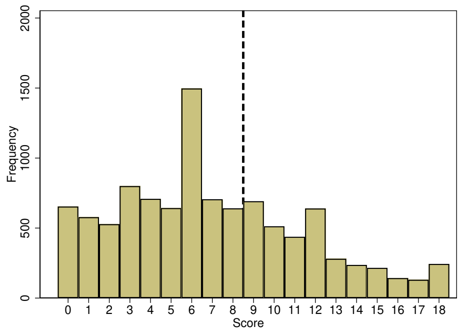
Figure 2.2: Histogram of Running Variable for RD Analysis by categories



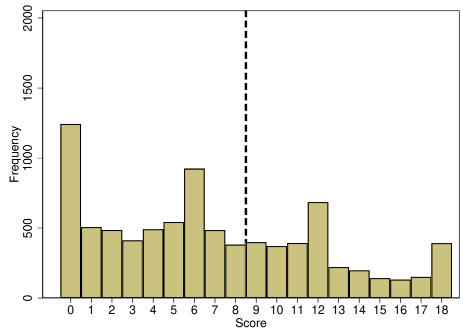
(a) Panel A: Z5 category



(b) Panel B: Z6 category



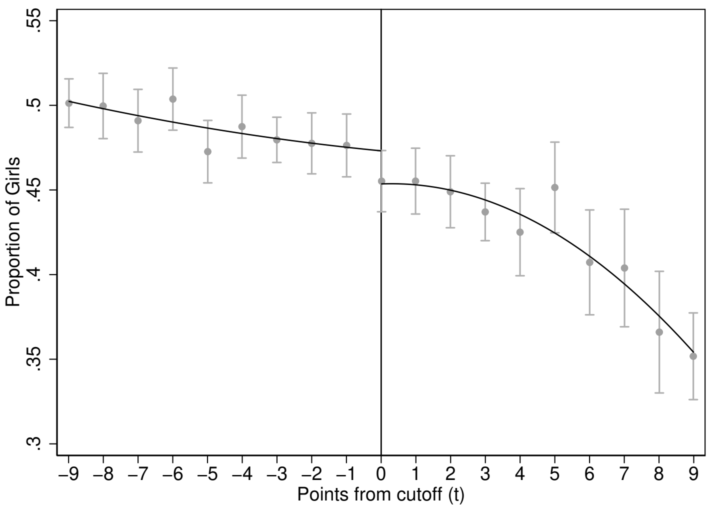
(c) Panel C: Z7 category



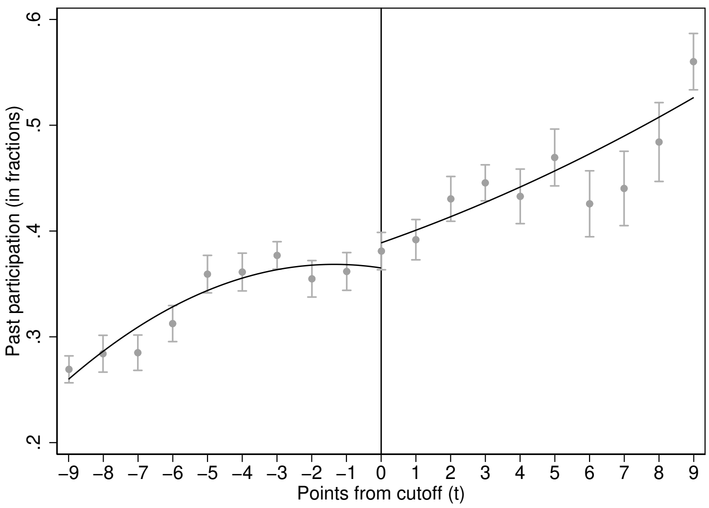
(d) Panel D: Z8 category

Notes: Samples is all student-test results for each category (Z5 to Z8) between 2011-2017 (as in Table 2.1, Panels B, C, D, and E). Bars reported within bins of width 1. Sample sizes for categories Z5, Z6, Z7, and Z8 are 16,334, 11,777, 10,291, and 8,566, respectively.

Figure 2.3: Predetermined Covariates by Points from Cutoff



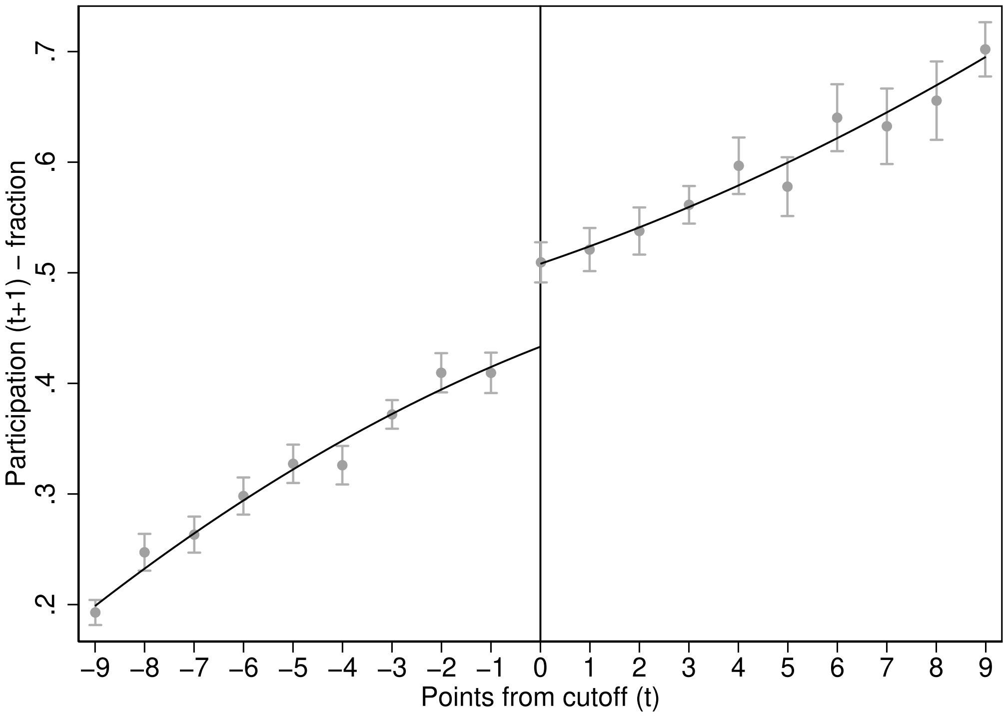
(a) Gender



(b) Past participation

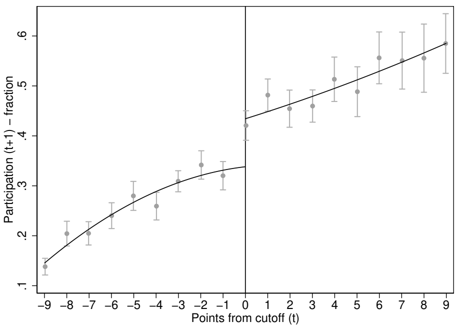
Notes: Vertical axis shows the average fraction of girls (Panel (a)), and fraction of students with previous experience (Panel (b)) who participated in the Olympiad. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received and $S_i \leq 0$ indicating negative feedback received. Mean and 95% confidence intervals are shown for predetermined covariates within each bin (score). The figure also shows the estimated polynomial in points margin allowing for a discontinuity at the 0 margin. Sample is all student-test results in categories Z5 to Z8 between 2011-2017 (as in Table 1, column 2). Sample size is 46,968.

Figure 2.4: Feedback Effect on Future Participation

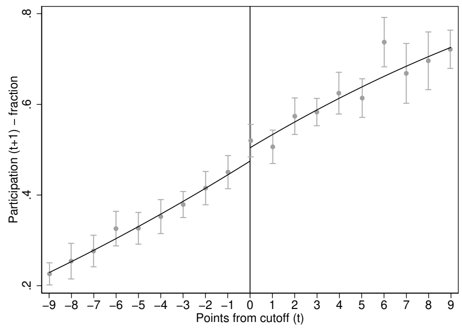


Notes: Vertical axis shows the fraction of students who participate in the Olympiads the year after. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received and $S_i \leq 0$ indicating negative feedback received. Mean and 95% confidence intervals are shown for the outcome within each bin (score). The figure also shows the estimated polynomial in points margin allowing for a discontinuity at the 0 margin. Sample is all student-test results in categories Z5 to Z8 between 2011-2017 (as in Table 2.1, column 2). Sample size is 46,968.

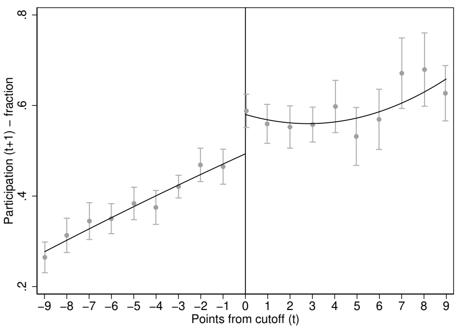
Figure 2.5: Feedback Effect on Future Participation by categories



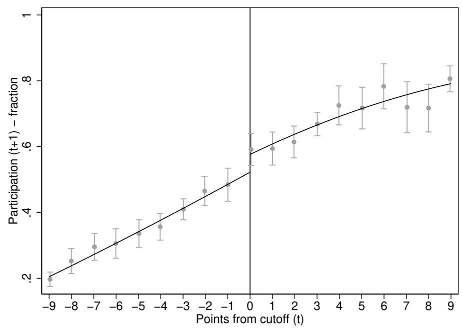
(a) Panel A: Z5 category



(b) Panel B: Z6 category



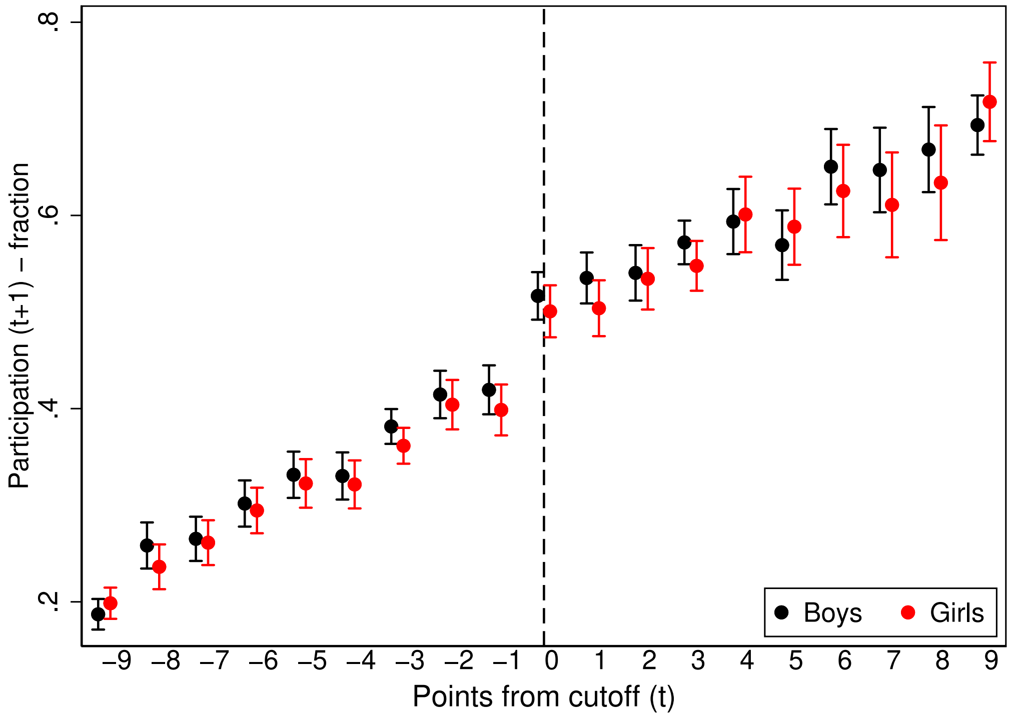
(c) Panel C: Z7 category



(d) Panel D: Z8 category

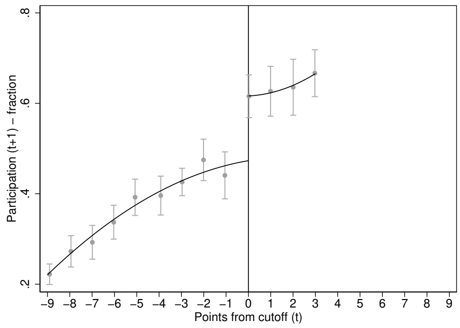
Notes: Vertical axis shows the fraction of students who participate in the Olympiads the following year by category. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received and $S_i \leq 0$ indicating negative feedback received. Mean and 95% confidence intervals are shown for the outcome within each bin (score). Sample sizes for categories Z5, Z6, Z7, and Z8 are 16,334, 11,777, 10,291, and 8,566, respectively.

Figure 2.6: Feedback Effect on Future Participation by Gender

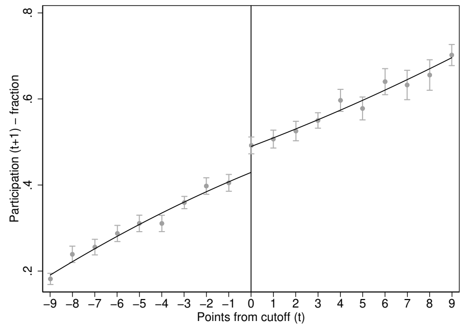


Notes: Vertical axis shows the fraction of students who participate in the Olympiads the following year by gender. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received and $S_i \leq 0$ indicating negative feedback received. Mean and 95% confidence intervals are shown for the outcome within each bin (score). Sample is all student-test results in categories Z5 to Z8 between 2011-2017 (as in Table 2.1, column 2). Sample size for boys and girls are 25,032 and 21,936, respectively.

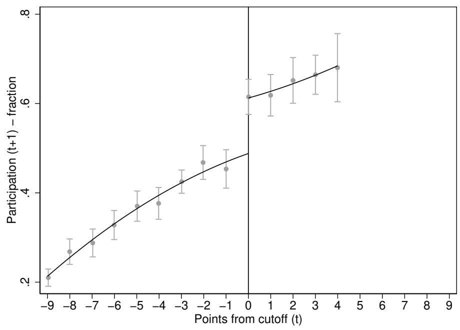
Figure 2.7: Feedback Effect on Future Participation by Maximum Score Achieved in Districts



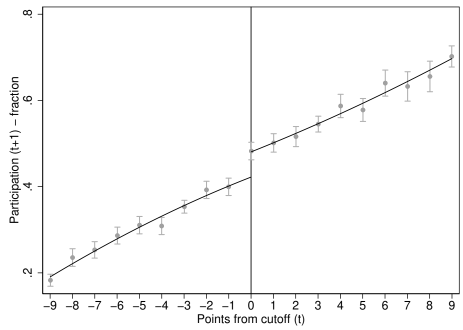
(a) Max district score< 13 (margin< 4)



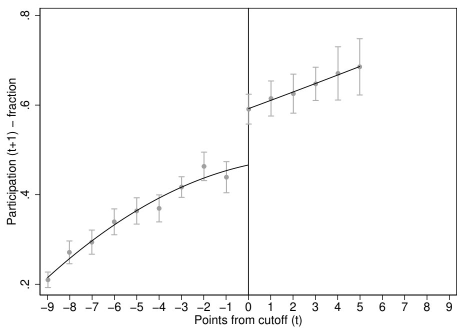
(b) Max district score≥ 13 (margin≥ 4)



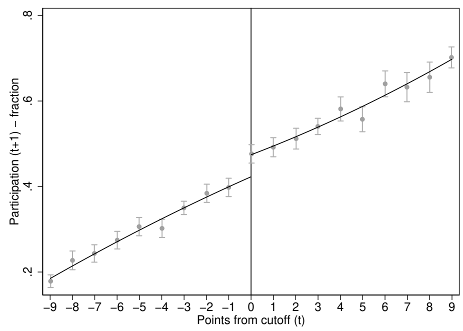
(c) Max district score< 14 (margin< 5)



(d) Max district score≥ 14 (margin≥ 5)



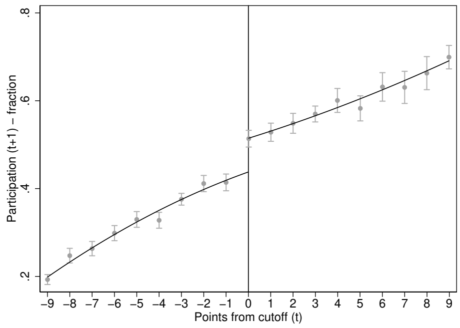
(e) Max district score< 15 (margin< 6)



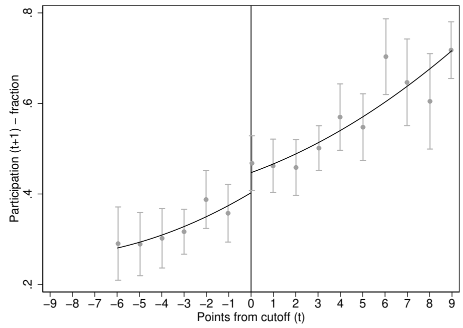
(f) Max district score≥ 15 (margin≥ 6)

Notes: Vertical axis shows the fraction of students who participate in the following category. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received. Mean and 95% confidence intervals are shown for the outcome within each bin (score).

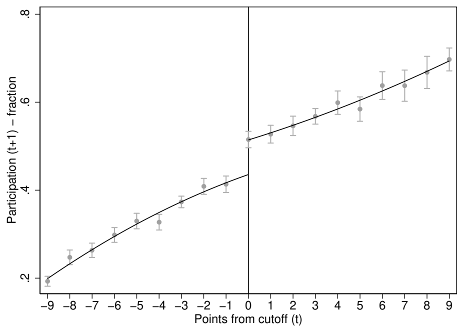
Figure 2.8: Feedback Effect on Future Participation by Minimum Score Achieved in Districts



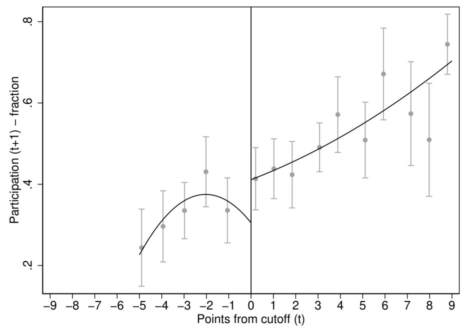
(a) Min district score< 3 (margin< -6)



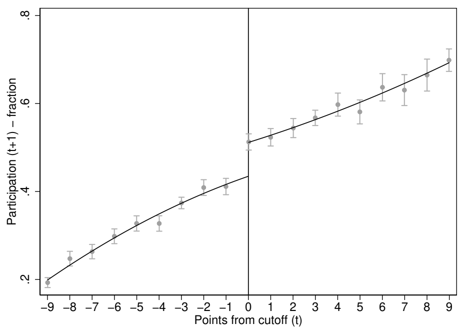
(b) Min district score≥ 3 (margin≥ -6)



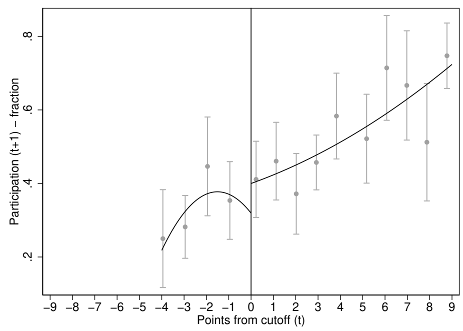
(c) Min district score< 4 (margin< -5)



(d) Min district score≥ 4 (margin≥ -5)



(e) Min district score< 5 (margin< -4)



(f) Min district score≥ 5 (margin≥ -4)

Notes: Vertical axis shows the fraction of students who participate in the Olympiads in the next category. Horizontal axis shows normalised score S_i to feedback cutoff, with $S_i \geq 0$ indicating positive feedback received. Mean and 95% confidence intervals are shown for the outcome within each bin (score).

Tables

Table 2.1: Sample Characteristics

		RD samples (points from cutoff)	
	All results (1)	+/-2 points (2)	+/-3 points (3)
Panel A. All transitions			
Past participation (percent)	37	37	38
Female (percent)	47	47	47
Hungarian (percent)	8	8	8
Grammar school (percent)	14	14	14
Observations	46,968	11,182	18,656
Panel B. Z5 to Z6			
Past participation (percent)	0	0	0
Female (percent)	47	47	47
Hungarian (percent)	8	7	8
Grammar school (percent)	5	5	5
Observations	16,334	4,115	6,653
Panel C. Z6 to Z7			
Past participation (percent)	46	48	47
Female (percent)	47	48	48
Hungarian (percent)	8	8	8
Grammar school (percent)	18	17	17
Observations	11,777	2,876	4,558
Panel D. Z7 to Z8			
Past participation (percent)	61	63	63
Female (percent)	47	46	47
Hungarian (percent)	8	8	9
Grammar school (percent)	19	19	19
Observations	10,291	2,552	4,487
Panel E. Z8 to Z9			
Past participation (percent)	66	71	71
Female (percent)	46	44	44
Hungarian (percent)	9	7	8
Grammar school (percent)	18	20	19
Observations	8,566	1,639	2,958

Notes: Sample in column 1 includes one observation per student-test (categories Z5 to Z8) between 2011 and 2017. Sub-samples in columns 2-4 include student-test observations whose scores fell within +/- 2 and 3 points from the cutoff, respectively.

Table 2.2: The Effect of Positive Feedback on Predetermined Characteristics of Competitors

	Gender		Past participation	
	RD sample		RD sample	
	+/-2pts (1)	+/-3pts (2)	+/-2 pts (3)	+/-3pts (4)
$1(S \geq 9)$	-0.0188	-0.0215**	0.00349	0.01676
	(0.0104)	(0.00760)	(0.00304)	(0.00856)
S	0.00137	-0.00155	0.01881***	0.00855**
	(0.00146)	(0.00324)	(0.00155)	(0.00147)
$S \times 1(S \geq 9)$	-0.0111	-0.00121	-0.01285*	0.00105
	(0.0154)	(0.00253)	(0.00406)	(0.00303)
θ_{dtc}	Yes	Yes	Yes	Yes
θ_s	Yes	Yes	Yes	Yes
Observations	10,912	18,522	10,912	18,522
R2	0.255	0.188	0.533	0.503

Notes: This table contains regression discontinuity based estimates of the effect of receiving positive feedback on predetermined characteristics for two different bandwidths. Gender is an indicator variable indicating whether the participant is a girl, and past participation is an indicator variable indicating whether the participant participated in a previous category. Parentheses contain standard errors clustered at the participant level.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.3: Effect of Positive Feedback on Subsequent Participation

	RD samples			
	+/-2 points		+/-3 points	
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	0.0526***	0.0337**	0.0388**	0.0388***
	(0.00805)	(0.0104)	(0.0121)	(0.00755)
$Score$	0.0145***	0.0253	0.0276***	0.0285***
	(0.00108)	(0.0117)	(0.00326)	(0.00440)
$Score \times 1(Score \geq 9)$	0.00826*	-0.0109	-0.0152***	-0.0145***
	(0.00329)	(0.00937)	(0.00349)	(0.00342)
Constant	0.346***	0.316**	0.392***	0.380***
	(0.0265)	(0.0543)	(0.0266)	(0.0295)
θ_d	Yes	No	Yes	No
θ_t	Yes	No	Yes	No
θ_c	Yes	No	Yes	No
θ_s	Yes	Yes	Yes	Yes
θ_{dtc}	No	Yes	No	Yes
Observations	11,054	10,912	18,567	18,522
Peer groups		1,553		1,714
R2	0.255	0.478	0.216	0.406

Notes: This table contains the estimates of receiving positive feedback on future participation based on regression discontinuity regression. Controls included in the regressions are gender of competitors, whether they participated before, and their rank within their district. Parentheses contain standard errors doubled clustered at the participant and points levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.4: The Effect of Positive Feedback on Subsequent Participation by Gender

	RD samples			
	+/-2 points		+/-3 points	
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	0.0539*	0.0167	0.0466**	0.0326***
	(0.0170)	(0.0102)	(0.0141)	(0.0079)
$Score$	0.0161*	0.0337**	0.0262***	0.0280***
	(0.0067)	(0.0091)	(0.0033)	(0.0039)
$Score \times 1(Score \geq 9)$	0.0052	-0.0329**	-0.0191***	-0.0195**
	(0.0074)	(0.0057)	(0.0033)	(0.0056)
$Girl$	-0.0209	-0.0491***	-0.0095*	-0.0214**
	(0.0216)	(0.0061)	(0.0046)	(0.0080)
$Girl \times 1(Score \geq 9)$	-0.0028	0.0355***	-0.0169*	0.0127
	(0.0204)	(0.0061)	(0.0081)	(0.0089)
$Girl \times Score$	-0.0034	-0.0167*	0.0029	0.0016
	(0.0139)	(0.0059)	(0.0022)	(0.0039)
$Girl \times S \times 1(S \geq 9)$	0.0066	0.0469**	0.0089*	0.0108
	(0.0204)	(0.0133)	(0.0038)	(0.0072)
Constant	0.347***	0.338***	0.389***	0.385***
	(0.0175)	(0.0479)	(0.0261)	(0.0281)
θ_d	Yes	No	Yes	No
θ_t	Yes	No	Yes	No
θ_c	Yes	No	Yes	No
θ_s	Yes	Yes	Yes	Yes
θ_{dtc}	No	Yes	No	Yes
Observations	11,054	10,912	18,567	18,522
Peer groups		1,553		1,714
R2	0.255	0.478	0.217	0.407

Notes: This table contains the estimates of the effect of receiving positive feedback on future participation by gender of participants based on the regression discontinuity design. All regressions control for past participation and rank within the district. Parentheses contain standard errors doubled clustered at the participant and points levels.

*** Significant at the 1 percent level

** Significant at the 5 percent level

* Significant at the 10 percent level

Table 2.5: Sample Characteristics - Upper Score Comparison

	Break point=13		Break point=14		Break point=15	
	< 13	≥ 13	< 14	≥ 14	< 15	≥ 15
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Past part. (percent)	46	35	43	35	41	35
Female (percent)	48	46	48	46	48	46
Hungarian (percent)	7	8	6	9	6	9
Grammar (percent)	11	14	10	15	10	15
Observations	7,290	39,678	10,276	36,692	13,976	32,992

Notes: This table describes competitors' characteristics under three different classifications of districts depending on the maximum score achieved. Columns (1) and (2) correspond to the partition based on a break point of 13 points. For instance, column (1) groups all competitors coming from districts where the maximum score obtained by a competitor was 12 or less, while column (2) groups all competitors coming from districts where the maximum score obtained by a competitor was 13 or more. Columns (3) and (4) correspond to the partition based on a break point of 14 points, while columns (5) and (6) correspond to the partition based on a break point of 15 points.

Table 2.6: Sample Characteristics - Lower Score Comparison

	Break point=3		Break point=4		Break point=5	
	< 3	≥ 3	< 4	≥ 4	< 5	≥ 5
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Past part. (percent)	36	48	36	48	37	50
Female (percent)	47	47	47	46	47	46
Hungarian (percent)	8	4	8	2	8	3
Grammar (percent)	14	15	14	15	14	16
Observations	43,640	3,328	45,027	1,941	45,883	1,085

Notes: This table describes competitors' characteristics under three different classifications of districts depending on the minimum score achieved. Columns (1) and (2) correspond to the partition based on a break point of 3 points. For instance, column (1) groups all competitors coming from districts where the minimum score obtained by a competitor was 2 or less, while column (2) groups all competitors coming from districts where the minimum score obtained by a competitor was 3 or more. Columns (3) and (4) correspond to the partition based on a break point of 4 points, while columns (5) and (6) correspond to the partition based on a break point of 5 points.

Table 2.7: Tests for Random Assignment of Maximum and Minimum Scores in District

	RD samples			
	+/-2 points		+/-3 points	
	Max	Min	Max	Min
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	-0.0652	0.0033	-0.0440	-0.0226
	(0.0389)	(0.0184)	(0.0406)	(0.0218)
$Score$	0.0927**	0.0283	0.0552**	0.0438***
	(0.0215)	(0.0121)	(0.0149)	(0.0060)
$Score \times 1(Score \geq 9)$	-0.0256	-0.0504*	0.0349	-0.0292*
	(0.0238)	(0.0176)	(0.0223)	(0.0120)
Observations	10,626	10,626	17,815	17,815
R2	0.587	0.566	0.564	0.522

Notes: This table contains the estimated effect of receiving positive feedback on the maximum and minimum scores achieved in a district. All regressions include district-by-year, category and school fixed effects. Parentheses contain standard errors doubled clustered at the participant and points levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.8: Positive Feedback and Score Comparison (Max and Min)

	RD samples			
	+/-2 points		+/-3 points	
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	0.227**	0.178**	0.220***	0.161***
	(0.041)	(0.043)	(0.032)	(0.032)
$Score$	0.019***	0.010**	0.029***	0.021***
	(0.001)	(0.003)	(0.003)	(0.002)
$Score \times 1(Score \geq 9)$	0.011*	0.002	-0.011**	-0.012***
	(0.004)	(0.003)	(0.004)	(0.003)
$Score^{max}$	-0.008	-0.005	-0.011***	-0.008**
	(0.004)	(0.004)	(0.003)	(0.003)
$Score^{min}$	0.013***	0.001	0.010***	-0.005
	(0.002)	(0.002)	(0.002)	(0.004)
$Score^{max} \times 1(Score \geq 9)$	-0.011**	-0.008**	-0.011***	-0.008***
	(0.003)	(0.003)	(0.002)	(0.002)
$Score^{min} \times 1(Score \geq 9)$	-0.000	0.005	-0.006	-0.002
	(0.005)	(0.004)	(0.006)	(0.007)
Constant	0.497***	0.346**	0.591***	0.441***
	(0.083)	(0.075)	(0.059)	(0.062)
θ_d	Yes	No	Yes	No
θ_t	Yes	No	Yes	No
θ_c	Yes	No	Yes	No
θ_s	Yes	Yes	Yes	Yes
θ_{dt}	No	Yes	No	Yes
Observations	10,665	10,656	17,845	17,842
R2	0.258	0.381	0.220	0.336

Notes: This table contains the estimated effect of receiving positive feedback on future participation depending on score comparisons. All regressions control for gender, past participation and rank within the district. Parentheses contain standard errors doubled clustered at the participant and points levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.9: Tests for Random Assignment of High- and Low-performing Peers (Definition 1)

	RD samples			
	+/-2 points		+/-3 points	
	High	Low	High	Low
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	-0.0008	-0.0001	0.0004	0.0040*
	(0.0006)	(0.0015)	(0.0007)	(0.0016)
Score	0.0022***	-0.0021*	0.0011***	-0.0055***
	(0.0004)	(0.0008)	(0.0003)	(0.0007)
$Score \times 1(Score \geq 9)$	-0.0012*	0.0007	-0.0002	0.0037***
	(0.0005)	(0.0009)	(0.0003)	(0.0009)
Observations	10,626	10,626	17,815	17,815
R2	0.565	0.582	0.538	0.543

Notes: This table contains the estimated effect of receiving positive feedback on the fraction of high- and low-performing peers in the district. High-performing peer is defined as the competitor who scores 18 points while a low-performing peer scores 0 points. All regressions include district-by-year, category and school fixed effects. Parentheses contain standard errors doubled clustered at the participant and points levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.10: Tests for Random Assignment of High- and Low-performing Peers (Definition 2)

	RD samples			
	+/-2 points		+/-3 points	
	High	Low	High	Low
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	-0.0000	0.0024	0.0009	0.0046**
	(0.0011)	(0.0015)	(0.0013)	(0.0012)
Score	0.0026**	-0.0055***	0.0011**	-0.0073***
	(0.0007)	(0.0007)	(0.0004)	(0.0005)
$Score \times 1(Score \geq 9)$	-0.0019*	0.0028	0.0012	0.0048***
	(0.0008)	(0.0012)	(0.0006)	(0.0008)
Observations	10,626	10,626	17,815	17,815
R2	0.583	0.578	0.559	0.544

Notes: This table contains the estimated effect of receiving positive feedback on the fraction of high- and low-performing peers in the district. High-performing peer is defined as the competitor who scores at least 17 points while a low-performing peer scores at most 1 point. All regressions include district-by-year, category and school fixed effects. Parentheses contain standard errors doubled clustered at the participant and points levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 2.11: Tests for Random Assignment of High- and Low-performing Peers (Definition 3)

	RD samples			
	+/-2 points		+/-3 points	
	High	Low	High	Low
	(1)	(2)	(3)	(4)
$1(Score \geq 9)$	-0.0007	-0.0016	0.0015	0.0010
	(0.0010)	(0.0019)	(0.0014)	(0.0015)
$Score$	0.0034**	-0.0054***	0.0013*	-0.0072***
	(0.0006)	(0.0008)	(0.0005)	(0.0006)
$Score \times 1(Score \geq 9)$	-0.0005	0.0029	0.0013*	0.0034**
	(0.0008)	(0.0015)	(0.0006)	(0.0010)
Observations	10,626	10,626	17,815	17,815
R2	0.576	0.577	0.558	0.539

Notes: This table contains the estimated effect of receiving positive feedback on the fraction of high- and low-performing peers in the district. High-performing peer is defined as the competitor who scores at least 16 points while a low-performing peer scores at most 2 points. All regressions include district-by-year, category and school fixed effects. Parentheses contain standard errors doubled clustered at the participant and points levels.

*** Significant at the 1 percent level

** Significant at the 5 percent level

* Significant at the 10 percent level

Table 2.12: Positive Feedback and Score Comparisons (High- and Low-performing Peers) - district-by-year fixed effects

	I: $H \geq 18pts, L \leq 0pts$		II: $H \geq 17pts, L \leq 1pt$		III: $H \geq 16pts, L \leq 2pts$	
	+/-2	+/-3	+/-2	+/-3	+/-2	+/-3
	(1)	(2)	(3)	(4)	(5)	(6)
$1(score \geq 9)$	0.039**	0.038**	0.041**	0.038**	0.042**	0.038**
	(0.010)	(0.011)	(0.010)	(0.011)	(0.009)	(0.011)
$score$	0.020***	0.027***	0.019***	0.026***	0.018***	0.026***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.004)
$1(score \geq 9) \times score$	0.003	-0.016***	0.003	-0.015***	0.005	-0.015***
	(0.007)	(0.003)	(0.006)	(0.003)	(0.006)	(0.004)
$High$	-0.002	-0.006	-0.018	-0.020**	-0.022	-0.026**
	(0.013)	(0.007)	(0.014)	(0.008)	(0.011)	(0.009)
Low	-0.021**	-0.026***	-0.039*	-0.036***	-0.050**	-0.041***
	(0.007)	(0.005)	(0.014)	(0.007)	(0.011)	(0.009)
$1(score \geq 9) \times High$	-0.023*	-0.027***	-0.023*	-0.026***	-0.018**	-0.022**
	(0.009)	(0.006)	(0.009)	(0.006)	(0.005)	(0.006)
$1(score \geq 9) \times Low$	-0.022**	-0.008	-0.012	-0.007	-0.001	-0.002
	(0.005)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Constant	0.332***	0.375***	0.326***	0.371***	0.309***	0.364***
	(0.038)	(0.027)	(0.042)	(0.029)	(0.033)	(0.034)
Observations	10,793	18,154	10,793	18,154	10,793	18,154
R2	0.322	0.269	0.324	0.271	0.324	0.271

Notes: This table contains the estimated effect of receiving positive feedback on future participation depending on score comparisons. All regressions control for gender, past participation and rank within the district. Parentheses contain standard errors doubled clustered at the participant and points levels. ***, **, * Significant at the 1, 5 and 10 percent level, respectively.

Table 2.13: Positive Feedback and Score Comparisons (High- and Low-performing Peers) - separate fixed effects

	I: $H \geq 18pts, L \leq 0pts$		II: $H \geq 17pts, L \leq 1pt$		III: $H \geq 16pts, L \leq 2pts$	
	+/-2	+/-3	+/-2	+/-3	+/-2	+/-3
	(1)	(2)	(3)	(4)	(5)	(6)
$1(score \geq 9)$	0.047***	0.038**	0.049***	0.038**	0.050***	0.039**
	(0.007)	(0.011)	(0.007)	(0.011)	(0.007)	(0.011)
$score$	0.012**	0.024***	0.012**	0.024***	0.011**	0.024***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
$1(score \geq 9) \times score$	0.009	-0.013**	0.010*	-0.012**	0.012**	-0.012**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$High$	0.012	0.008	-0.001	-0.001	-0.009	-0.009
	(0.012)	(0.007)	(0.012)	(0.007)	(0.013)	(0.008)
Low	-0.014	-0.021**	-0.025*	-0.024***	-0.037**	-0.031***
	(0.006)	(0.006)	(0.008)	(0.005)	(0.009)	(0.006)
$1(score \geq 9) \times High$	-0.023**	-0.026***	-0.023**	-0.026***	-0.020**	-0.023**
	(0.006)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)
$1(score \geq 9) \times Low$	-0.023**	-0.005	-0.014*	-0.004	-0.005	0.001
	(0.006)	(0.010)	(0.005)	(0.006)	(0.005)	(0.006)
Constant	0.321***	0.360***	0.319***	0.363***	0.309***	0.358***
	(0.044)	(0.026)	(0.044)	(0.027)	(0.042)	(0.027)
Observations	10,821	18,158	10,821	18,158	10,821	18,158
R2	0.257	0.218	0.257	0.218	0.258	0.219

Notes: This table contains the estimated effect of receiving positive feedback on future participation depending on score comparisons. All regressions control for gender, past participation and rank within the district. Parentheses contain standard errors doubled clustered at the participant and points levels. ***, **, * Significant at the 1, 5 and 10 percent level, respectively.

Chapter 3

Ordinal Rank Effects in Math Competitions

I study the impact of students' ordinal rank in district math competitions on their subsequent participation and performance by using administrative data of math Olympiad results in Slovakia. The analysis takes advantage of this setting where students are informed about their rank within districts. To estimate the impact of rank, I exploit idiosyncratic variation in the distribution of test scores across districts and cohorts. I find that students with a higher rank are more likely to participate the following year. Moreover, conditional on subsequent participation, students with a higher rank obtain higher scores the following year. In exploring mechanisms, I investigate whether rank effects operate through school choices. I take advantage of an institutional feature by which competitors can switch from regular to grammar schools at the end of the 5^{th} *grade. I find a positive but non-significant effect of rank on the probability to switch to grammar schools.*

3.1 Introduction

Recent papers have documented the importance of a student's rank during primary school for their later educational outcomes. Between two otherwise identical students in a class, the pupil with the higher rank performs better in secondary school (Yu, 2020), is more likely to finish high school and to attend college (Elsner and Isphording, 2017), and earns higher levels of income 19 years later (Denning, Murphy and Weinhardt, 2021). How does the individual's rank within a peer group impact their educational outcomes? The main mechanism boils down to the idea that higher ranked students perceive themselves as more intelligent which might lower their cost of effort (Murphy and Weinhardt, 2020). In the Psychology literature, the relation between rank and academic self-concept is known as the big-fish-little-pond effect which predicts that a student with a given ability will have a higher academic self-concept when surrounded by low-achieving peers than by high-achieving ones (Marsh et al., 2008). Thus, the importance of rank over individuals' outcomes ultimately hinges on the extent to which individuals compare against each other.

In this sense, attributing individuals' actions to their rank poses two important empirical challenges related to the difficulty of individuals to compare against each other. First, it is often difficult to define an appropriate peer group. For instance, Bandiera and Rasul (2006) find that individuals can have multiple reference groups to form expectations and make decisions. Second, and more importantly, it requires that members of a peer group be informed about each other's abilities, or able to extract such information, for social comparison to take place. To the best of my knowledge, in none of these studies are individuals informed about the abilities of their peers. As a consequence, estimations of rank effects must rely on the strong assumption that the rank among members of a peer group is learned by constant interaction (Elsner and Isphording, 2017) raising concerns about measurement error and whether estimates are capturing the effect of ordinal rank.

Using administrative data of mathematical Olympiads participants from 5 levels (5^{th} to 9^{th} grade) in Slovakia covering over 46,000 results, I study the effect of students' rank within a peer group on their subsequent outcomes at the competitions. In contrast to previous studies, two features of the mathematical Olympiads provide a unique opportunity to address this research question. First, competitors are grouped based on their geographical location and, therefore, the boundaries of reference groups are precisely defined. Second, a competitor is not only informed about her performance but also on that of her peers within the district. In Math Olympiads, district committees manage national examinations locally, and report the total score distribution in the district to participants, and their parents and teachers. By linking individual data of competitors over time, I observe whether they participate in the competition the following year, and their subsequent performance conditional on participating. I use the ordinal rank within a district known to all agents to test whether it affects these subsequent outcomes in Math Olympiads.

To identify a causal effect of ordinal rank, I study competitors with the same math ability (as measured by their achievement in the nation-wide mathematical Olympiads) who, by chance, are ranked differently in their districts. More formally, to hold own and peer ability constant, I follow Murphy and Weinhardt (2020) and leverage the idiosyncratic variation in the distribution of test scores across districts, category and cohorts. In this way, I compare students in different districts who have the same test score relative to their district mean, but different ranks due to differences in the shape of score distributions in their districts.

The main finding of this study is that a student's ordinal rank within a district has a strong impact on math Olympiad outcomes the year after. First of all, rank affects the likelihood of subsequent participation in the competition. Students ranked at the top of the district compared to the bottom are 17 percentage points more likely to compete again the next year. Second, conditional on subsequent participation, a higher ranked participant performs better in the next category of the mathematical Olympiads. Ranking at the top of class compared to the bottom is associated with a gain of 6.3 national percentiles. This short-run effect is slightly lower than the estimates in Murphy and Weinhardt (2020), who find an analogous figure of 7.9 national percentiles.

While most of the studies point out that rank effects are explained by students' reactions to their relative position, the rank effect can also include reactions of their parents or teachers. Indeed, Kinsler and Pavan (2021) find that parents of children attending schools with low average skills tend to believe their child is higher in the overall skills distribution than they actually are. Nonetheless, it is not clear how these distorted beliefs should translate into actions that would influence children's outcomes. Parents might further encourage their children to strengthen their mathematical skills, or they could compensate and focus on other dimensions of their children's upbringing (e.g. music and sports) while tutors might devote either lower or higher training to higher ranked students depending on their educational objectives. The way these individuals react to a child's higher rank depends on their preferences and, therefore, the direction of the effect is ambiguous *a priori*. In the existing literature there is a lack of evidence on parental reactions, which might be due to the fact that parents are poorly informed about their child's rank (Murphy and Weinhardt, 2020).

I investigate whether ordinal rank effects operate via competitors' school choices, in which parents play an important role. Since school choices are an important input of skill development, parents' reactions to their children's rank might partially explain the reduced-form rank effects on performance if higher ranks lead to more investment in the education of their children. To address this reaction, I study a crucial school choice: students who finish the 5^{th} grade have the option to change from a regular to a grammar school that will significantly improve their math training. The higher quality of these schools is observed in the data as, conditional on previous score, students who move to grammar schools obtain 6 national percentiles more in the subsequent Olympiad than students who stay in regular schools. Thus, I test whether students with a higher rank in the 5^{th} grade math Olympiad are more likely to move to a grammar school the year after.

I find a positive but non-significant rank effect on the probability of switching from regular to grammar schools. Students ranked at the top of the district compared to the bottom are 6 percentage points more likely to change from regular to grammar schools. Imprecision of results might be explained by sample size issues, as the analysis is restricted to 5^{th} grade competitors in the mathematical Olympiad.

This paper contributes to two strands of the literature. First, it extends the novel literature on ordinal rank effects on educational outcomes. Rank affects subsequent performance in secondary school (see Cicala, Fryer and Spenkuch (2018) and Yu (2020)), educational attainment (see Elsner and Isphording (2017)), choices in college (see Delaney and Devereux (2021)), future earnings (see Denning, Murphy and Weinhardt (2021)), and violent behaviour (see Comi et al. (2021)). This paper is similar to those listed above in terms of the identification strategy. However, this study makes one fundamental innovation by studying ordinal rank effects in a setting where individuals are perfectly informed about their ranks. In contrast, the previous literature is based on settings where individuals are imperfectly informed or non-informed at all about their relative position[1]. As a result, rank estimates might be not only driven by social comparison forces but also by competitive traits of a selected sample of students who are concerned about their rank and do their best to extract such information.

This paper also relates to the literature on the support for high-achieving

[1]In Elsner and Isphording (2017), rank measure is based on results of IQ tests that are not observable to students of primary schools. In Murphy and Weinhardt (2020), primary school students are given only one of five broad attainment levels, with 85% of students achieving one of the top two levels. In Denning, Murphy and Weinhardt (2021), the authors state "students will not necessarily be told their class rank in these exams by their teachers, nor do we believe that students care particularly about their ranking in these low-stakes examinations". In Elsner, Isphording and Zölitz (2021), rank of university students is not communicated.

students in mathematics. There is evidence on the role played by high schools (Ellison and Swanson, 2016), socioeconomic background (see Hoxby and Avery (2013) and Agarwal and Gaule (2020*b*)), and gender (see Ellison and Swanson (2010*b*) and Iriberri and Rey-Biel (2019)). This paper contributes to this literature by studying a novel determinant of the development of gifted students. The ordinal rank in mathematical competitions, which are a gateway through which gifted students develop their skills, proves to be an important factor in encouraging children to participate and improve their performance in mathematics.

In terms of policy, the main finding of this paper has a practical implication: local competitions should inform participants on their rank at the national level rather than at the district level. By doing so, competitors who are "small fishes" in very competitive districts would form non-distorted beliefs about their abilities and remain interested in competing. Such a policy will erase asymmetries of information that might be leading to sub-optimal decisions of competitors.

The remainder of this paper is as follows. Section 3.2 describes the institutional setting of competitions and the educational system. Section 3.3 presents the data. Section 3.4 shows the empirical analysis. Section 3.5 explores mechanisms. Section 3.6 concludes.

3.2 Institutional Setting

In this section I describe the organisation of mathematical Olympiads in Slovakia and explain why this setting is suitable for studying ordinal rank effects. I also explain the educational system, as competitors are subject to school choices during the period of study. This institutional feature is important since I investigate school choices as a potential channel of rank effects.

3.2.1 Math Olympiads in Slovakia

Mathematical Olympiads in Slovakia are yearly national competitions for elementary (5^{th} to 9^{th} grades) and secondary school (10^{th} to 13^{th} grades) students, and are organised by the Slovak Committee of Mathematical Olympiad (SKMO). For each grade, the SKMO designs a national test consisting of 3 or 4 proof-based questions. However, the structure of the competitions depends on the students' grade: 1) from 5^{th} to 8^{th} grade, students only compete in a district round; 2) from 9^{th} to 11^{th} grade, an additional regional round is set for around half of the best students from the district round; 3) for 12th and 13th grade, a national round is added on top of the district and regional ones, and is meant for a small and selected group of best students at the regional round. Figure 3.1 summarises the structure of Mathematical Olympiads in Slovakia.

This analysis focuses on competitions designed for 5th to 9th grade students, for which data at the district level is available. Tests for 5th to 8th grade students consist on 3 questions (6 points each), and students who score at least 9 points are recognised as successful, while the 9th grade test involves 4 questions (6 points each), and the same recognition is given at 12 points. Olympiads are categorised by adding the prefix "Z" to the correspondent grade. For instance, Olympiads for 5th and 9th grade students are called "Z5" and "Z9" categories, respectively.

To conduct Mathematical Olympiads at the district level, the SKMO forms committees in each of the 78 districts to manage these competitions locally. The district committee is responsible for testing, marking examinations following a uniform scoring criteria, and publishing results. Reports are produced for each grade at the district level, handed to teachers in each school that sent competitors within the district, and also published in the SKMO website (free access). These reports contain full names of participants, names of their schools, scores obtained, and the ranking within the district. Appendixes A.3 and A.4 contain examples of these reports for a given district in categories Z7 in 2016 and Z8 in 2017, respectively.

The way the SKMO disseminates Olympiads' results is crucial for the purpose of this study. A competitor (and her tutors and parents) is not only informed about her score, but also her rank and the score/rank of all other competitors in the category in her district. It is often difficult to define an appropriate peer group (Dahl, Løken and Mogstad, 2014). Although competitors in a given district know the whole score distribution in their district, they ignore the national score distribution[2]. To illustrate, consider competitors "A" and "B" who obtained the same score at a given category in the same year but a different rank in their respective district such that "A" is among the students with the lowest score in his district, while "B" is among the students with highest score in his district. Due to incomplete information about the na-

[2]For a given district, the report containing results of all competitors (see Appendixes A.3 and A.4) is handed to tutors from schools who sent competitors. Competitors/tutors/parents can query the report in the SKMO website as well. However, in order to know the national score distribution for a given category and year, one must download and process each of the 78 district reports (in PDF format) available at the SKMO website.

tional score distribution, competitor “B” might have the wrong belief of being one of the brightest students in the competition. I provide a formal explanation on how to isolate the effect of ordinal rank on subsequent outcomes at the mathematical Olympiads based on the score distribution differences across district-cohorts in section 3.4 .

3.2.2 School system in Slovakia

Subjects of this study belong to the school level called “lower secondary” education in Slovakia which comprises 5^{th} (10 years old) to 9^{th} (14 years old) grade students. Throughout this level, students can enroll in three types of schools that focus on skills in different ways. First, *základné školy* (ZŠ), or regular school, follows the national curriculum and provide uniform education to students. Second, *gymnázium*, or grammar school, provides academic education for very talented students, and admission is subject to very competitive examinations. Finally, conservatories teach pupils the general curriculum with an emphasis on music classes to become either professional musicians or teachers. Figure 3.2 shows how these three different types of school coexist in “lower secondary” levels, and how mathematical Olympiads fit into it.

As shown in Figure 3.2, all students must be enrolled in a regular school at 5^{th} grade. Only after finishing this grade do students have the option to move to a grammar school (8-year *gymnázium* program) or conservatory (8-year conservatory program). Students who decided to remain in regular schools have again the option to move to a grammar school (5-year *gymnázium* program) after finishing the 8^{th} grade. Finally, after finishing the 9^{th} grade, all students who remain in regular schools conclude the “lower secondary” level, and must choose between three different tracks for the “upper secondary” level: grammar school (4-year *gymnázium* program), secondary vocational school (2, 3, 4 or 5-year program), and conservatory (6-year program).

In this regard, I focus on the optional change from regular to grammar schools by the end of the 5^{th} grade. Although grammar schools do not imply higher tuition fees in comparison to regular schools, switching to grammar schools might be costly, as students need to prepare for highly competitive admission tests and also travel farther since grammar schools are less available than regular schools. This is more problematic for students living in small towns or villages where grammar schools are absent. It is worth mentioning that most parents are willing to switch their children to grammar schools as it represents a crucial educational investment: grammar schools have better teachers, deliver knowledge in greater depth, and impose higher study requirements (Federičová and Munich, 2014). I provide suggestive evidence of the better mathematical training provided in grammar schools in comparison to regular schools conditional on prior mathematical abilities (as measured by scores obtained in Mathematical Olympiad) in Section 3.5.1. Finally, I do not focus on switching to conservatories as these schools do not emphasize academic education but rather music skills. In fact, no student from conservatories was found in reports of district competitions.

3.3 Data

3.3.1 Sample description

I collected data for all students participating at district Mathematical Olympiads in categories Z5 to Z9 between 2011 and 2018. The data set is based on 2,375 district reports in PDF format (see Appendixes A.3 and A.4 for examples). Individual data includes full names, the test taken (Z5 to Z9), the score, the rank within the district, and the school affiliation. In addition, I inferred the sex of participants based on their names[3], the language of instruction (Slovak or Hungarian) based on the school description[4], and the type of school (regular or grammar) based on their official description[5].

To obtain panel data of competitors, I built unique identifiers for competitors based on names provided and consistency over time. As shown in Appendixes A.3 and A.4, reports present participants' names in two different ways: either first name/last name or last name/first name. Thus, three tasks were necessary to conduct in order to identify competitors over time. First, I identified first names and last names by using *Facebook* and *Namepedia* websites. Second, by using text analysis algorithms, I calculated a similarity index of any possible pair of names to identify names with small variation due to

[3]I identified the sex of competitors based mainly in their first names by using Facebook and Namepedia. Moreover, I also take advantage on the fact that some last names also indicate the sex of individuals. In particular, Slovak last names that terminate in "ová" or "á" are female last names. I use this information to corroborate the sex of participants inferred from their first names.

[4]I extracted the language of instruction of schools from their website. In particular, on top of indicating the type of school, they also specify whether the language of instruction is Hungarian by adding *vyučovacím jazykom maďarským* (meaning "language of instruction is Hungarian") next to official school name.

[5]As shown in Appendixes A.3 and A.4, reports present the name of schools with preceded by either "*gymnázium*" or "*ZŠ*" standing for *základné školy* which means regular schools. The type of school was corroborated by the official description in its website.

typos such as missing one character (e.g., "s" instead of "š") which could correspond to the same individual. Finally, I assigned individual identifiers based on time consistency. In particular, a given ID should be consistent with years and categories. For example, competitor with name "X" participating in category Z6 held in 2013 should match with competitor with name "X" in category Z7 held in 2014 in the same district but no with competitor with name "X" participating in category Z8 held in year 2014.

As the interest of this paper is to investigate the effect of students' ordinal rank at time t on outcomes in the next category at time $t+1$, I built one-year transitions for categories Z5 to Z8. For instance, the transition for the Z5 category includes all competitors who took the Z5 category during the period 2011-2017 and their outcomes in the Z6 category during the period 2012-2018. In total, the data contains 46,968 individual transitions for each category (Z5 to Z8) in the period 2011-2017 to subsequent categories (Z6 to Z9) in the period 2012-2018. Given this structure, it is important to note that a given student can appear up to 4 times in this data (Z5 to Z6, Z6 to Z7, Z7 to Z8, and Z8 to Z9).

Test scores at each category-year are converted into a national test score percentile to make them comparable across years. Since 5$^{\text{th}}$ to 8$^{\text{th}}$ grade tests are marked in a discrete scale from 0 to 18 points, there are 19 transformed scores in a range from 5 to 100. For the 9$^{\text{th}}$ grade test, the maximum number of points is 24 and, therefore, there are 25 transformed scores in a range from 5 to 100.

Table 3.1 shows descriptive statistics for participants in each of the four categories. Regarding similarities in competitors' characteristics among categories, Table 3.1 shows that test scores have a mean of around 50 with standard deviation of around 30 in all categories. This is due to the fact that test scores were converted to percentiles. Participation of girls and individuals whose language of instruction is Hungarian are also stable across categories. Girls and Hungarian students account for around 47% and 8% of total participants in each category, respectively.

Regarding differences in participants' characteristics across categories, Table 3.1 shows that the average past participation varies depending on the category. Naturally, there are no participants with previous experience at the Olympiads at Z5 category (5$^{\text{th}}$ grade) since tests are not designed for lower grades. As we move forward, the average past participation increases. In the Z8 category, 66% of competitors have previous experience in at least one of the categories Z5 to Z7. Likewise, future participation also increases over categories. While 33% of participants in the Z5 category compete again in the Z6 category, 45% of participants in the Z8 category compete again in the Z9 category.

3.3.2 Ordinal Rank

The regressor of interest is a student's ordinal rank among her district peers. I use the rank at the district level because this is where score/rank comparison takes place. As explained in section 3.2.1, this ordinal ranking is formally reported to all competitors within their district, as opposed to all settings studied in this literature where ranking is not informed and it must be assumed that students have some perception of their rank based on interactions with peers along with repeated teacher feedback.

Thus, I use the reported raw position p_{icd} of competitor i of category c in district d of size N_{cd} (for examples, see Appendixes A.3 and A.4). For example, the best student within a district has $p = 1$, while the worst one has $p = N_{cd}$. Since district cohorts vary in size, this raw rank (e.g., 1, 2, and 3) is not comparable across districts of different sizes. Therefore, I convert the raw rank to a percentile rank r_{icd} using the following formula:

$$r_{icd} = \frac{N_{cd} - p_{icd}}{N_{cd} - 1} \tag{3.1}$$

By construction, this rank r_{icd} is uniformly distributed and bounded between 0 and 1, with the lowest-ranked student in each district-category having $r = 0$ and the highest ranked having $r = 1$. For instance, the worst competitor shown in Appendix A.3 corresponding to a district of 25 competitors ($p_{icd} = 25, N_{cd} = 25$) has $r_{icd} = 0$, and so does the worst competitor shown in Appendix A.4 corresponding to a district of 28 competitors ($p_{icd} = 28, N_{cd} = 28$). If two or more competitors obtain the same score, the committee gives them the higher rank.

Table 3.2 presents descriptive statistics of the transformed ordinal rank and districts' size for each category. The ordinal rank has a mean of 0.55 with a standard deviation of around 0.30 for all categories. It is important to note that although the raw ordinal rank was converted to a percentile rank, the latter has a mean larger than 0.5. This is due to the fact that in the case of ties, both students are given the higher rank. Regarding the number of competitors at the district level, Table 3.2 shows that the average number of competitors per district decreases over categories. For instance, in the Z5 category, the average number of competitors per district is 43.23, while in the Z8 category, it is 26.20. This table also reveals that the number of competitors varies considerably across districts. For instance, in the category Z5, its standard deviation is 20.07, and the minimum and maximum number of competitors in a district are 3 and 118, respectively.

3.3.3 Outcome variables

The panel structure of this data allows for the tracking of students who participated in any of the 4 categories (Z5 to Z8) during 2011-2017, and the observation of two subsequent outcomes the year after in the correspondent categories (Z6 to Z9) during 2012-2018. First, I observe competitors' participation the year after in the following category. For example, the data shows whether a given competitor in the Z5 category in 2011 shows up for the Z6 category in 2012. This subsequent participation is the first outcome variable (extensive margin). Second, conditional on attending the following category the next year, the data contains this subsequent performance. For instance, for competitors in the Z5 category in 2011 who also attended the Z6 category in 2012, the score obtained in the latter is the second outcome variable (intensive margin).

I also construct an intermediate outcome variable (the school choice after 5^{th} grade) as follows. For students who participated in the Z5 category during 2011-2017 (who are enrolled in regular schools by law as shown in Figure 3.2), and also attended in any competition afterwards (Z6 to Z9), I observe whether they remain in the same regular school or change to a grammar school based on the school description. For instance, if a Z5 competitor in 2011 showed up in the Z6 category in 2012, this outcome variable records whether the student remains in a regular school.

3.4 Empirical Analysis

3.4.1 Identification Strategy

In order to isolate the effect of ordinal rank, I follow the novel and most rigorous approach developed by Murphy and Weinhardt (2020) to hold own and peer ability constant. Before explaining the formal approach, I intuitively describe the identification strategy by using examples drawn from the data set. Figure 3.3 shows the test score distributions in 3 districts corresponding to the Z5 category held in 2016. Each district has the same minimum (0 points), maximum (18 points), and mean (8.2 points) score in the Z5 category test. Furthermore, in each district there is a competitor scoring 12 points. However, due to differences in the shape of score distributions, each of these competitors has a different rank within the corresponding district. In these examples, the ranks of the competitors scoring 12 points range between 0.71 and 0.93, despite having the same absolute and relative to-the-district-mean scores. This considerable variation will allow us to identify the effect of rank conditional on district fixed effects which will remove all observable and non-observable mean differences between districts.

The previous example only considered one category in a given year. Since our data set comprises multiple categories and years, I will exploit the differences in the test score distributions across districts, categories, and years. More formally, to account for factors that would impact all competitors within a district for a given category and year (e.g., any peer group characteristic such as mean ability or the variance in ability, the presence of a disruptive peer, tutors, or unexpected events on the test day), I include fixed effects at the district-category-year level following Murphy and Weinhardt (2020). While this transformation centres the (residual) score distribution of each district-category-year at the same mean, it does not modify the shape of the score distribution. Therefore, despite controlling for district-category-year fixed effects, the ordinal ranking is preserved and is identified from differences across districts, categories and years in higher moments of the score distribution (Elsner, Isphording and Zölitz, 2021).

3.4.2 Model Specification

To estimate the impact of ordinal rank in district on a competitor's subsequent participation and performance, I use the following equation:

$$y_{id,c+1,t+1} = \gamma r_{idct} + g(s_{idct}) + X_i'\beta + \lambda_{dct} + \theta_s + \epsilon_{id,c+1,t+1}, \tag{3.2}$$

where the dependent variable $y_{id,c+1,t+1}$ is one of the following two outcomes at the mathematical Olympiads: 1) indicator variable for participation of competitor i in district d in subsequent category $c+1$ held in year $t+1$, or 2) her correspondent achievement conditional on participation. I regress these outcomes on r_{idct} which is the percentile ranking of competitor i in district d in category c held in year t. To isolate the rank effect for a given score, I control for a function of s_{idct} which is the absolute score (in percentile) of competitor i in district d in category c held in year t. X_i controls for predetermined individual characteristics including gender and language of instruction (Slovak or Hungarian). Crucial for identification, I follow Murphy and Weinhardt (2020) by conditioning on district-category-year fixed effects λ_{dct}. Finally, θ_s is a school fixed effect. It is important to emphasise that competitors under analysis participated in at least one of the categories Z5 to Z8 during years 2011-2017, that is $c = Z5, ..., Z8$; and $t = 2011, 2012, .., 2017$.

Regarding the relationship between future outcomes in Olympiads and prior scores, Denning, Murphy and Weinhardt (2021) indicate that incorrect functional forms of g might cause specification error and create a spurious rank effect. For the sake of robustness, I present results using four different choices of $g()$: linear and nonlinear functions (19 indicators for all but the tenth ventiles of competitors' score), and second and third-degree polynomials.

3.4.3 Rank Effects on Outcomes

Tables 3.3 and 3.4 display the estimated effects of ordinal rank on subsequent participation and performance. These tables report coefficients from separate regressions using different functional forms for the relationship between subsequent outcomes and prior score. Column 1 includes score as a linear function, column 2 includes score in a nonlinear fashion (19 indicators for all but the tenth ventiles of competitors' score), column 3 adjusts for a second-order polynomial in score, and column 4 conditions on a third-order polynomial in score. All regressions control for individual characteristics (gender and past participation), and the fixed effects described in section 3.4.2. Each coefficient shows the marginal effect of an increase in a student's ordinal rank within a district, holding constant individual score and mean peer score as well as all other factors that are constant across all competitors of a district-category-year. It can be seen that for the first outcome analysis (subsequent participation), there are 46,922 observations from 1,787 districts-category-year groups. However, for the second outcome analysis (subsequent performance), there are 19,071 observations from 1,387 districts-category-year groups. The difference in observations and groups is due to the fact that this second analysis is restricted to competitors who participated in the next category.

Columns 1 to 4 in Table 3.3 show the rank impact on subsequent participation in the mathematical Olympiad. The interpretation of the estimate from column 1 is that ranking at the top of the district compared to the bottom increases the likelihood of participating in the next category by 17 percentage points. Columns 2 to 4 reveal that the magnitude, precision, and significance of the coefficients are stable across specifications. Columns 1 to 4 in Table 3.4 show the rank impact on subsequent performance conditional on participating in the mathematical Olympiad. The interpretation of the estimate from column 1 is that ranking at the top of the district compared to the bottom is associated with a gain of 5.74 national percentiles. The choice of different functional forms for prior scores seems to affect the point estimates' magnitude. For instance, when allowing for a cubic function in prior scores (column 4), the associated gain is 6.29 national percentiles. However, given the standard errors of the estimated effects, we cannot reject the hypothesis that these effects are different across different functional forms.

It is important to mention that these estimates are based on students who show up the next year. To have an idea on how this affects estimates, Elsner and Isphording (2017) assumed that these students would have obtained zero points if they had participated.

3.5 School Choices

In this section I attempt to investigate channels through which the reduced-form rank effect operates. In this regard, rank effects might reflect not only students' reactions, but also those of others around them, such as parents and teachers (Denning, Murphy and Weinhardt, 2021). While there is suggestive evidence that students with higher ranks perceive themselves as more intelligent (which might explain the observed rank effects), the literature is less informative about the role played by students' parents. The lack of evidence on these studies might be due to the fact that parents are poorly informed about their children' rank (Murphy and Weinhardt, 2020). In fact, in none of those settings are parents informed of students' ranks. Only by interacting with teachers might they know at most whether their child is above or below the mean in a peer group.

To fill this gap in the literature, I study whether ordinal rank effects operate via educational investments where parents play an important role. It is worth noting that the way parents react to a child's higher rank is an empirical question as the theoretical direction of the effect is ambiguous: parents might further encourage their children to strengthen their math skills, or they could compensate and focus on other dimensions of their children's upbringing (e.g. leisure and sports). To address the former reaction, I study school choices made by the end of the 5$^{\text{th}}$ grade. As explained in Section 3.2.2, the Slovak educational system gives the option to change from a regular to a grammar school to students who finish 5$^{\text{th}}$ grade. This represents an important educational investment since grammar schools have better teachers, deliver knowledge in greater depth, and impose higher study requirements (Federičová and Munich, 2014).

To address school choices as a potential external mechanism of the rank effect, I test whether students with a higher rank in the 5$^{\text{th}}$ grade math Olympiad are more likely to move to a grammar school the year after. Before presenting those results, I provide evidence that switching from regular to grammar schools improves competitors' mathematical training and, therefore, serves as a potential mechanism of the rank effect.

3.5.1 Mathematical Training in Grammar Schools

To explore whether switching to grammar schools improves competitors' mathematical training, I compare the performance of competitors who switch to grammar schools versus competitors who remain in regular schools holding constant their prior mathematical performance (as measure by their score in Z5 category). It is important to note that this is not a causal analysis of the effect of grammar schools on competitors' performance. Although conditioning on prior scores at the mathematical Olympiad provides a credible way to hold constant their innate mathematical abilities (Agarwal and Gaule, 2020*b*), the data set does not provide information on competitors' characteristics that determine the decision to switch from regular to grammar schools (e.g., performance in subjects other than mathematics, parental background characteristics, etc.). Moreover, this analysis is based only on competitors who participate in the Z5 category and then in any other further category (Z6 to Z9) as this is the only way I could observe whether, after attending a regular school, they transfer to a grammar school.

Figure 3.4 sheds light on the positive relationship between the decision to switch to a grammar school and subsequent performance at the mathematical Olympiad. Panel (a) shows the average performance in the Z6 category of competitors who switch to grammar schools versus competitors who remain in regular schools conditional on their score in the Z5 category (in quintiles). For each quintile, the average performance of competitors who switch to grammar schools is significantly higher than that of competitors who remain in regular schools. This means that switching to a grammar school increases the performance of competitors regardless of their level of prior ability. Panel (b), (c), and (d) shows the same analysis for the Z5 category competitors who further participate at Z7, Z8, and Z9 categories. Although the subsequent average performance has larger confidence intervals due to smaller samples, the positive relationship between grammar school transfer and performance remains for all quintiles.

Next, I test this relationship using a multivariate analysis. Table 3.5 presents the results of regressing competitors' subsequent performance on an dummy variable indicating whether they switch to grammar schools conditional on performance in the Z5 category, and district, year, and school fixed effects. Columns (1) to (4) correspond to separate regressions based on samples used in Figure 3.4. For instance, column (1) shows the effect of switching to a grammar school on subsequent performance in the Z6 category which was depicted in Figure 3.4, panel (a). Column (5) gives the average result for all categories by stacking samples used in columns (1) to (4).

Table 3.5 reveals a highly significant relationship between attending a grammar school and performance at the mathematical Olympiads. For instance, competitors who switch to grammar schools obtain 6.9 national percentiles more in the Z6 category than students who stay in regular schools. This positive association is also observed in categories Z7 to Z9. However, columns (2) to (4) seem to suggest that the effect is lower as we move forward in the categories. For instance, for Z5 category competitors who also compete in the Z9 category, the effect of switching to grammar schools is 3.3 national percentiles and non-significant. This might be due to two reasons. The first reason is that imprecision might be caused by small samples. Indeed, only 1,358 competitors in the Z5 category (out of 16,334 as shown in Table 3.1) are observed in the Z9 category. The second reason might be that unobserved competitors' characteristics become more salient. We could expect that competitors who switch to grammar schools are more engaged in attending mathematical Olympiads than competitors who remain in regular schools. In this sense, competitors who stay in regular schools and continue attending the competitions for the following 4 years (in categories Z6 to Z9) might possess different levels of non-cognitive traits that make them different from the competitors who remain in regular schools and drop out of the competitions.

These results taken together suggest that grammar schools provide better mathematical training for mathematical Olympiads. Furthermore, anecdotal evidence given by members of the SKMO indicates that parents are aware of the higher quality and more challenging education imparted in grammar schools. Therefore, the question is whether competitors who are higher ranked in the Z5 category are more likely to switch to grammar schools, which might partially explain the reduced-form rank effects on subsequent performance.

3.5.2 Rank Effects on School Choices

I estimate the impact of ordinal rank at district level in the Z5 category on a competitor's likelihood to switch to a grammar school afterwards using a specification similar to equation (3.2):

$$y_{id,t+1} = \gamma r_{idt} + g(s_{idt}) + X_i\beta + \lambda_{dt} + \theta_s + \epsilon_{id,t+1}, \tag{3.3}$$

where the dependent variable $y_{id,t+1}$ is a dummy variable indicating whether competitor i in district d switches from a regular to a grammar school in year $t+1$. I regress this outcome on r_{idt} which is the percentile ranking of competitor i in district d in the Z5 category held in year t. To isolate the rank effect for a given score, I control for a function of s_{idt} which is the absolute score (in percentile) of competitor i in district d at the Z5 category held in year t. X_i controls for predetermined individual characteristics including gender and language of instruction (Slovak or Hungarian). Similar to equation (3.2), I condition on district-year fixed effects λ_{dt}. Finally, θ_s is a school fixed effect.

The implication of the institutional design on school choices described in section 3.2.2 is that the analysis is restricted to competitors who participated in the Z5 category during years 2011-2017 and at least in one of the subsequent categories Z6, Z7, Z8, and Z9 during years 2012-2018. Therefore, the only difference with respect to equation (3.2) is that the outcome (switching to a grammar school by the end of the 5$^{\text{th}}$ grade) is category-independent, which explains why there is no category index in the specification. Similar to the main analysis of rank effects on subsequent outcomes, I estimate rank effects on school choices using different functional forms for the relationship between school choices and score in the Z5 category.

Results are reported in Table 3.6. I run six separate regressions depending on the choice function to model the relationship between school choices and score in the Z5 category mathematical Olympiad. Column (1) reports the rank effect on the likelihood of switching to a grammar school using a linear function of the Z5 category score. Columns (2), (3), and (4) use quintiles, deciles and ventiles, respectively. Columns (5) and (6) use second-order and third-order polynomials. Table 3.6 shows a positive but non-significant rank effect on the probability of switching from regular to grammar schools for all specification, except for column (2) which conditions on quintiles of the Z5 category score. Imprecision of results might be explained by a considerable reduction in sample size. Indeed, there are only 6,373 participants in the Z5 category who are also observed in any subsequent category.

3.6 Conclusions

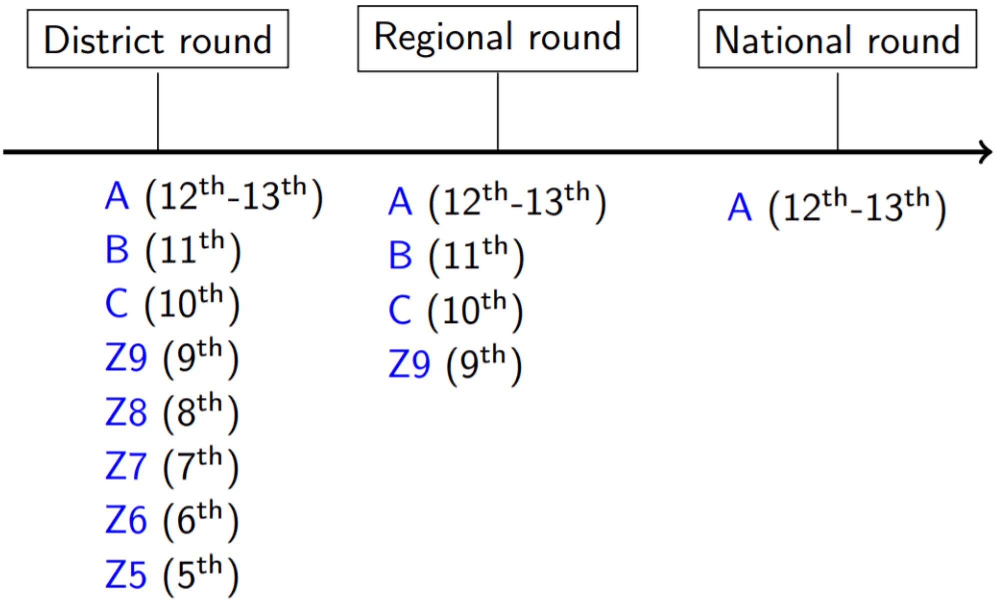
In this paper I study ordinal rank effects on educational outcomes in the short-run. I conduct this study in the context of mathematical Olympiads designed for children from the 5^{th} to 9^{th} grades. More importantly, in this setting all individuals (competitors, their tutors and parents) are perfectly informed about their absolute performance and rank within their district. There are two main findings. First, a higher ordinal rank in a district increases the probability of attending the competition the year after. Second, conditional on participating again, a higher-ranked student obtains a higher score.

I inspect whether the rank effect on outcomes at the mathematical Olympiads might be explained by school choices that provide better training to competitors. By taking advantage of an institutional setting in Slovakia in which students can switch from a regular to a grammar school by the end of the 5^{th} grade, I study whether higher-ranked competitors are more likely to switch to grammar schools. I find positive but non-significant effects likely related to the sample size issues.

In terms of policy implications, this paper shows that the way tests' results are delivered to individuals might lead to inefficient decisions. In particular, since competitors only observe the score distribution in their district, they might form distorted beliefs (in comparison to a global comparison) about their mathematical abilities and make sub-optimal decisions on subsequent participation and training for the Olympiad. The policy recommendation to solve such inefficiency is to inform competitors about their rank at the national level instead of at the district level. By doing so, competitors who are "small fishes" in very competitive districts would form precise beliefs about their abilities and potentially remain interested in competing and training.

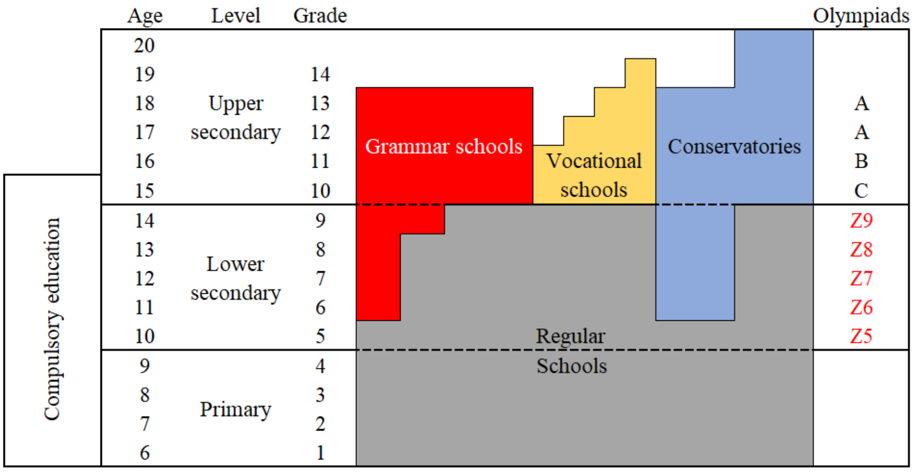
Figures

Figure 3.1: Math Olympiads Structure



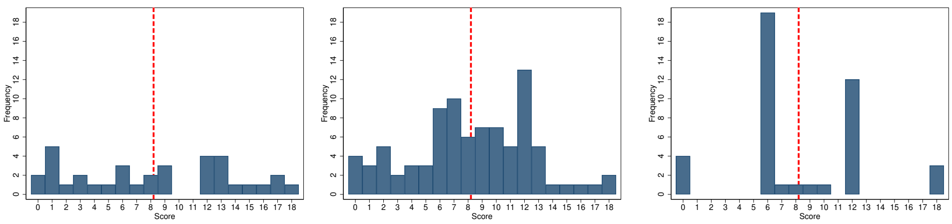
Notes: This figure shows the structure of Mathematical Olympiads depending on the grade assessed. There are three rounds available (district, regional, and national) depending on the grade.

Figure 3.2: Slovak School System and Math Olympiads



Notes: This figure shows the school system in Slovakia and how Mathematical Olympiads are designed for each grade. Description includes the correspondent age, level, grades and the different types of school that coexist for different grades.

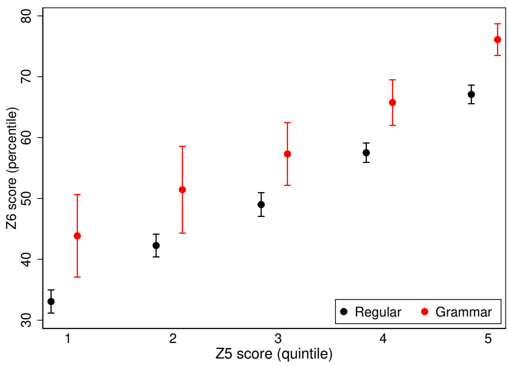
Figure 3.3: Variation in Rank



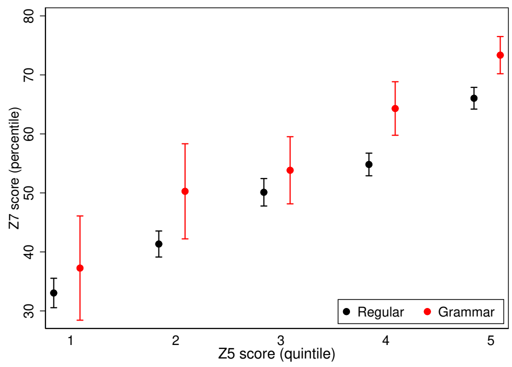
(a) R(12)=0.71 (b) R(12)=0.87 (c) R(12)=0.93

Notes: This figure illustrates the score distributions (category Z5, year 2015) from three districts in the data set. All three districts have a mean score of 8.2 (shown in red dashed line) and the same minimum and maximum scores (0 and 18, respectively). Dependent on the district's exact shape of the score distribution, a competitor scoring 12 points may end up with different ranks between the 71^{th} and 93^{th} percentile.

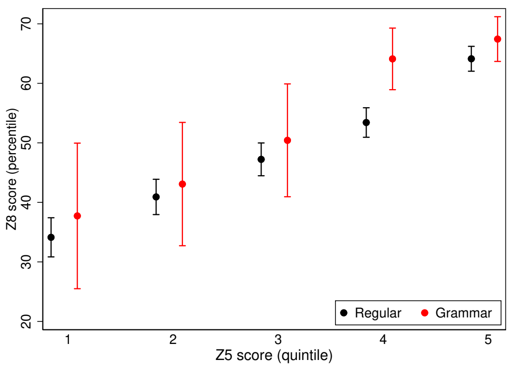
Figure 3.4: Subsequent Performance in Regular vs Grammar Schools



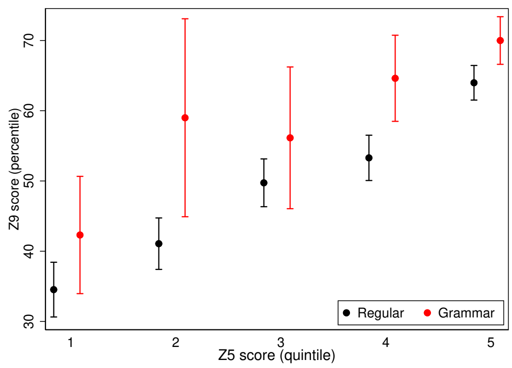
(a) Z5 to Z6



(b) Z5 to Z7



(c) Z5 to Z8



(d) Z5 to Z9

Notes: The x-axis shows the score in the Z5 test (quintile). The y-axis shows the average of the score in later tests (percentile). Panels (a), (b), (c) and (d) show results for Z6, Z7, Z8, and Z9 tests. All students belonged to a regular school in the Z5 test. The year after, the student either remains in the same school (label "regular") or moves to a selective school (label "grammar").

Tables

Table 3.1: Sample Characteristics

	N (1)	Mean (2)	SD (3)	Min (4)	Max (5)
Panel A: Z5					
National score percentile	16,334	48.83	29.70	5	100
Past part. (previous category)	16,334	0	0	0	1
Future part. (next category)	16,334	0.33	0.47	0	1
Female	16,334	0.47	0.50	0	1
Hungarian	16,334	0.08	0.26	0	1
Grammar	16,334	0.05	0.21	0	1
Panel B: Z6					
National score percentile	11,777	48.84	29.48	5	100
Past part. (previous category)	11,777	0.46	0.50	0	1
Future part. (next category)	11,777	0.45	0.50	0	1
Female	11,777	0.47	0.50	0	1
Hungarian	11,777	0.08	0.28	0	1
Grammar	11,777	0.18	0.39	0	1
Panel C. Z7					
National score percentile	10,291	48.77	29.30	5	100
Past part. (previous category)	10,291	0.61	0.49	0	1
Future part. (next category)	10,291	0.45	0.50	0	1
Female	10,291	0.47	0.50	0	1
Hungarian	10,291	0.08	0.28	0	1
Grammar	10,291	0.19	0.39	0	1
Panel D. Z8					
National score percentile	8,566	47.85	29.87	5	100
Past part. (previous category)	8,566	0.66	0.47	0	1
Future part. (next category)	8,566	0.45	0.50	0	1
Female	8,566	0.46	0.50	0	1
Hungarian	8,566	0.09	0.28	0	1
Grammar	8,566	0.18	0.39	0	1

Notes: This table describes the characteristics of participants in each category. For each category, there is only one result for each participant.

Table 3.2: Ordinal Rank Statistics

	Mean (1)	SD (2)	Min (3)	Max (4)
Panel A: Z5				
Ordinal Rank	0.55	0.29	0	1
Number of competitors (district)	43.23	20.07	3	118
Panel A: Z6				
Ordinal Rank	0.55	0.29	0	1
Number of competitors (district)	34.80	16.32	2	92
Panel A: Z7				
Ordinal Rank	0.55	0.29	0	1
Number of competitors (district)	31.34	16.20	1	90
Panel A: Z8				
Ordinal Rank	0.56	0.29	0	1
Number of competitors (district)	26.20	12.73	1	63

Notes: This table presents rank statistics of participants within a district in each category. For each district-category, there is only one result for each participant.

Table 3.3: The Impact of Rank on Subsequent Participation

	Subsequent Participation			
	(1)	(2)	(3)	(4)
Rank	0.17***	0.17***	0.18***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)
Score function				
Linear	Yes	No	No	No
Nonlinear (ventiles)	No	Yes	No	No
2º polynomial	No	No	Yes	No
3º polynomial	No	No	No	Yes
Observations	46,922	46,922	46,922	46,922
Peer groups	1,787	1,787	1,787	1,787
R2	0.363	0.364	0.363	0.363

Notes: This table presents the estimates of the rank effect on subsequent participation at the mathematical Olympiad. I run separate regressions depending on the function choice to model the relationship between the subsequent outcomes and the previous score in the Olympiads. All regressions control for gender and past participation. A peer group is defined as a group of competitors in a given district from a given category held at a particular year. Standard errors clustered at the competitor and district levels are shown in parentheses.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 3.4: The Impact of Rank on Subsequent Performance

	Subsequent Performance			
	(1)	(2)	(3)	(4)
Rank	5.74**	6.21**	6.34***	6.29***
	(2.38)	(2.44)	(2.41)	(2.42)
Score function				
Linear	Yes	No	No	No
Nonlinear (ventiles)	No	Yes	No	No
2º polynomial	No	No	Yes	No
3º polynomial	No	No	No	Yes
Observations	19,071	19,071	19,071	19,071
Peer groups	1,367	1,367	1,367	1,367
R2	0.385	0.386	0.386	0.387

Notes: This table presents the estimates of the rank effect on subsequent performance at the mathematical Olympiad. I run separate regressions depending on the function choice to model the relationship between the subsequent outcomes and the previous score in the Olympiads. All regressions control for gender and past participation. A peer group is defined as a group of competitors in a given district from a given category held at a particular year. Standard errors clustered at the competitor and district levels are shown in parentheses.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 3.5: Grammar School Effects on Subsequent Performance

	Z5 to Z6 (1)	Z5 to Z7 (2)	Z5 to Z8 (3)	Z5 to Z9 (4)	All (5)
Grammar	6.957***	5.014**	5.455*	3.280	5.954***
	(1.607)	(2.004)	(2.930)	(2.072)	(1.395)
Obs	4,954	3,410	2,241	1,358	12,671
R2	0.363	0.367	0.399	0.475	0.299

Notes: This table presents the estimates of grammar school effects on subsequent performance at the mathematical Olympiad. Columns (1) to (4) show separate regressions for individuals who participated in Z5 and also in Z6, Z7, Z8, and Z9, respectively. Column (5) considers all samples used for calculation in columns (1) to (4). All regressions control for score in the Z5 category (percentile) and include district, year and school fixed effects. Column (5) includes in addition category fixed effects. In columns (1) to (4), standard errors are clustered at the district levels. In column (5) standard errors are clustered at the competitor and district levels.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Table 3.6: The Impact of Rank on School Choices

	Switching from Regular to Grammar School					
	(1)	(2)	(3)	(4)	(5)	(6)
Rank	0.041	0.134**	0.057	0.060	0.062	0.057
	(0.067)	(0.060)	(0.065)	(0.065)	(0.065)	(0.064)
Score function						
Linear	Yes	No	No	No	No	No
Quintiles	No	Yes	No	No	No	No
Deciles	No	No	Yes	No	No	No
Ventiles	No	No	No	Yes	No	No
2^o polynomial	No	No	No	No	Yes	No
3^o polynomial	No	No	No	No	No	Yes
Observations	6,373	6,373	6,373	6,373	6,373	6,373
Peer groups	483	483	483	483	483	483
R2	0.426	0.427	0.429	0.430	0.428	0.428

Notes: This table presents the estimates of the rank effect on the likelihood of switching from regular to grammar schools. I run separate regressions depending on the choice function to model the relationship between the school choices and the score in the Z5 category mathematical Olympiad. All regressions control for gender of competitors and include district-year and school fixed effects. A peer group is defined as a group of competitors in a given district from the Z5 category held at a particular year. Standard errors clustered at the district levels are shown in parentheses.
*** Significant at the 1 percent level
** Significant at the 5 percent level
* Significant at the 10 percent level

Bibliography

Agarwal, Ruchir and Patrick Gaule. 2020*a*. "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights* 2(4):409–424.

Agarwal, Ruchir and Patrick Gaule. 2020*b*. "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights* 2(4):409–424.

Arnold, David, Will Dobbie and Crystal S Yang. 2018. "Racial Bias in Bail Decisions." *The Quarterly Journal of Economics* 133(4):1885–1932.

Ash, Elliott, Daniel L. Chen and Arianna Ornaghi. 2020. "Gender Attitudes in the Judiciary: Evidence from U.S. Circuit Courts." *Working Paper* .

Azmat, Ghazala and Nagore Iriberri. 2010. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics* 94(7-8):797–811.

Bagues, Manuel and Berta Steve-Volart. 2010. "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment." *The Review of Economic Studies* 77(4):1301–1328.

Bagues, Manuel, Mauro Sylos-Labini and Natalia Zinovyeva. 2017. "Does the Gender Composition of Scientific Committees Matter?" *American Economic Review* 107(4):1207–1238.

Bandiera, Oriana and Imran Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *The Economic Journal* 116:869–902.

Bandiera, Oriana, Valentino Larcinese and Imran Rasul. 2015. "Blissful ignorance? A natural experiment on the effect of feedback on students' performance." *Labour Economics* 34:13–25.

Bedard, Kelly, Jacqueline Dodd and Shelly Lundberg. 2021. "Can Positive Feedback Encourage Female and Minority Undergraduates into Economics?" *AEA Papers and Proceedings* 111:128–132.

Boyd, Christina L. 2016. "Representation on the Courts? The Effects of Trial Judges' Sex and Race." *Political Research Quarterly* 69(4):788–799.

Boyd, Christina L., Lee Epstein and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54(2):389–411.

Buser, Thomas and Huaiping Yuan. 2019. "Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad." *American Economic Journal: Applied Economics* 11(3):225–252.

Card, David, Alexandre Mas, Enrico Moretti and Emmanuel Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102(6):2981–3003.

Chiappori, Pierre-André and Yoram Weiss. 2007. "Divorce, Remarriage, and Child Support." *Journal of Labor Economics* 25(1):37–74.

Cianci, Anna M., Howard J. Klein and Gerard H. Seijts. 2010. "The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness." *Journal of Applied Psychology* 95(4):618–630.

Cicala, Steve, Roland G. Fryer and Jörg L. Spenkuch. 2018. "Self-Selection and Comparative Advantage in Social Interactions." *Journal of the European Economic Association* 16(4):983–1020.

Collins, Todd and Laura Moyer. 2008. "Gender, Race, and Intersectionality on the Federal Appellate Bench." *Political Research Quarterly* 61(2):219–227.

Comi, Simona, Federica Origo, Laura Pagani and Marco Tonello. 2021. "Last and furious: Relative position and school violence." *Journal of Economic Behavior and Organization* 188:736–756.

Coontz, Phyllis. 2000. "Gender and judicial decisions: Do female judges decide cases differently than male judges?" *Gender Issues* 18(4):59–73.

Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2):448–474.

Dahl, Gordon B., Katrine V. Løken and Magne Mogstad. 2014. "Peer Effects in Program Participation." *American Economic Review* 104(7):2049–2074.

Delaney, Judith M. and Paul J. Devereux. 2021. "High School Rank in Math and English and the Gender Gap in STEM." *Labour Economics* 69.

Denning, Jeffrey T., Richard Murphy and Felix Weinhardt. 2021. "Class Rank and Long-Run Outcomes." *NBER Working Paper Series 27468* .

Dijk, Dina Van and Avraham Kluger. 2011. "Task type as a moderator of positive/negative feedback effects on motivation and performance: A regulatory focus perspective." *Journal of Organisational Behaviour* 32:1084–1105.

Dobrescu, Loretti Isabella, Marco Faravelli, Rigissa Megalokonomou and Alberto Motta. 2019. "Rank Incentives and Social Learning: Evidence from a Randomized Controlled Trial." *IZA Discussion Paper Series No. 12437* .

Ellison, Glenn and Ashley Swanson. 2010*a*. "The gender gap in secondary school mathematics at high achievement levels: Evidence from the American mathematics competitions." *Journal of Economic Perspectives* 24(2):109–128.

Ellison, Glenn and Ashley Swanson. 2010*b*. "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions." *Journal of Economic Perspectives* 24(2):109–128.

Ellison, Glenn and Ashley Swanson. 2016. "Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions." *American Economic Review* 106(6):1244–1277.

Ellison, Glenn and Ashley Swanson. 2021. "Dynamics of the Gender Gap in High Math Achievement." *The Journal of Human Resources* Forthcoming.

Elsner, Benjamin and Ingo E. Isphording. 2017. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics* 35(3):787–828.

Elsner, Benjamin, Ingo E. Isphording and Ulf Zölitz. 2021. "Achievement Rank Affects Performance and Major Choices in College." Forthcoming.

Farhang, Sean and Gregory Wawro. 2004. "Institutional Dynamics on the U.S. Court of Appeals: Minority Representation Under Panel Decision Making." *The Journal of Law, Economics, and Organization* 20(2):299–330.

Federičová, Miroslava and Daniel Munich. 2014. "The Impact of High-Stakes School-Admission Exams on Study Effort and Achievements: Quasi-experimental Evidence from Slovakia." *Journal of Population Economics* 30(3):1–24.

Gill, Rebecca, Michael Kagan and Fatma Marouf. 2015. "Chivalry, Masculinity, and the Importance of Maleness to Judicial Decision Making.".

Goulas, Sofoklis and Rigissa Megalokonomou. 2021. "Knowing who you actually are: The effect of feedback on short- and longer-term outcomes." *Journal of Economic Behavior and Organization* 183:559–615.

Gruhl, John, Cassia Spohn and Susan Welch. 1981. "Women as Policymakers: The Case of Trial Judges." *American Journal of Political Science* 25(2):308–322.

Hahn, Jinyong, Petra Todd and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69(1):201–209.

Hamermesh, Daniel S. 2001. "The Changing Distribution of Job Satisfaction." *Journal of Human Resources* 36(1):2981–3003.

Hoogveld, Nicky and Nick Zubanov. 2017. "The power of (no) recognition: Experimental evidence from the university classroom." *Journal of Behavioral and Experimental Economics* 67:75–84.

Hoxby, Caroline and Christopher Avery. 2013. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low-Income Students." *Brookings Papers on Economic Activity* (Spring):1–66.

Iriberri, Nagore and Pedro Rey-Biel. 2019. "Competitive Pressure Widens the Gender Gap in Performance: Evidence from a Two-stage Competition in Mathematics." *The Economic Journal* 129(620):1863–1893.

Kinsler, Josh and Ronni Pavan. 2021. "Local Distortions in Parental Beliefs over Child Skill." *Journal of Political Economy* 129(1):81–100.

Kiser, Randall L., Martin A. Asher and Blakeley B. McShane. 2008. "Let's Not Make a Deal: An Empirical Study of Decision Making in Unsuccessful Settlement Negotiations." *Journal of Empirical Legal Studies* 5(3):551–591.

Knepper, Matthew. 2018. "When the Shadow Is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases." *Journal of Labor Economics* 36(3):623–664.

Marsh, Herbert, Marjorie Seaton, Ulrich Trautwein, Oliver Ludtke, K. T. Hau, Alison J. O'Mara and Rhonda Craven. 2008. "The Big Fish Little Pond Effect Stands Up to Critical Scrutiny: Implications for Theory, Methodology, and Future Research." *Educational Psychology Review* 20(3):319–350.

Martin, Elaine. 1989. "Differences in Men and Women Judges: Perspectives on Gender." *Journal of Political Science* 17(1).

Mike, Anissa, Kelci Harris, Brent Roberts and Joshua Jackson. 2015. "Conscientiousness." *International Encyclopedia of the Social and Behavioral Sciences (Second Edition)* pp. 658–665.

Mouganie, Pierre and Yaojing Wang. 2020. "High-Performing Peers and Female STEM Choices in School." *Journal of Labor Economics* 38(3):805–841.

Murphy, Richard and Felix Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *Review of Economic Studies* 87(6):2777–2826.

Niederle, Muriel and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122(3):1067–1101.

Niederle, Muriel and Lise Vesterlund. 2010. "Explaining the Gender Gap in Math Test Scores: The Role of Competition." *Journal of Economic Perspectives* 24(2):129–144.

Peresie, Jennifer L. 2005. "Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts." *The Yale Law Journal* 114(7):1759–1790.

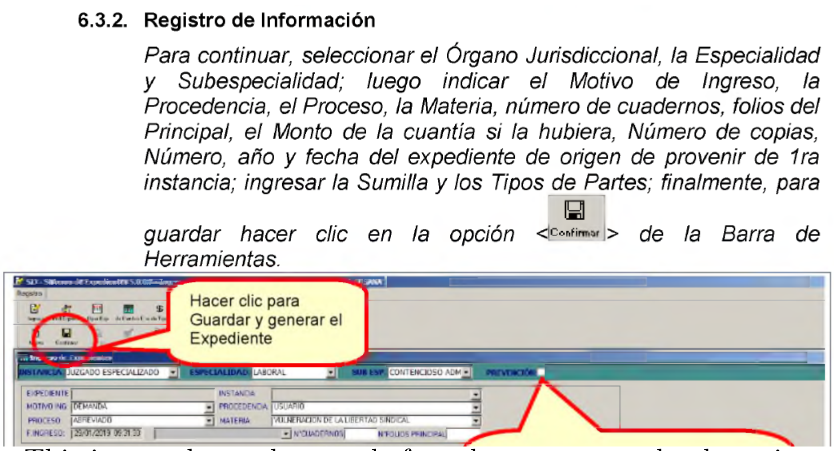
Sadka, Joyce, Enrique Vera and Christopher Woodruff. 2018. "Information and Bargaining through Agents: Experimental Evidence from Mexico's Labor Courts." *NBER Working Paper 25137* .

Villeval, Marie Claire. 2020. "Feedback policies and peer effects at work." *Handbook of Labor, Human Resources and Population Economics* .

Yu, Han. 2020. "Am I the big fish? The effect of ordinal rank on student academic performance in middle school." *Journal of Economic Behavior and Organization* 176:18–41.

Appendix A

A.1 Randomisation of cases in the Judicial System (before)

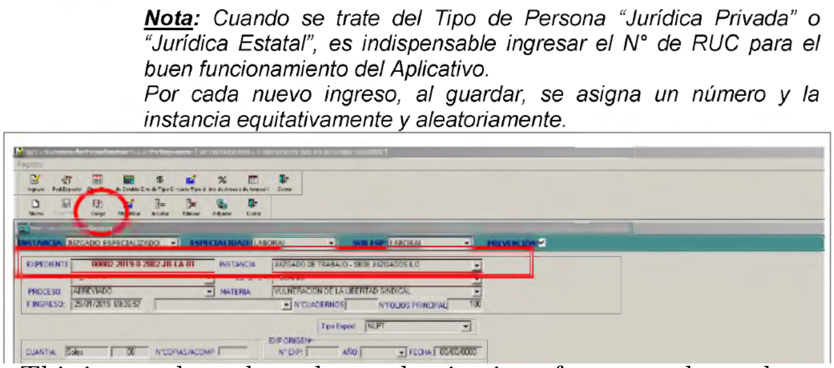


6.3.2. Registro de Información

Para continuar, seleccionar el Órgano Jurisdiccional, la Especialidad y Subespecialidad; luego indicar el Motivo de Ingreso, la Procedencia, el Proceso, la Materia, número de cuadernos, folios del Principal, el Monto de la cuantía si la hubiera, Número de copias, Número, año y fecha del expediente de origen de provenir de 1ra instancia; ingresar la Sumilla y los Tipos de Partes; finalmente, para guardar hacer clic en la opción <Confirmar> de la Barra de Herramientas.

Notes: This image shows the step before the system randomly assigns a case to a court within a district as shown in the user guide. As can be seen, the user, when registering the case into the system, cannot choose the court nor the ID of the case. Short translation: "To continue, select categories and characteristics of the case, then click on the save button".

A.2 Randomisation of cases in the Judicial System (after)



Notes: This image shows how the randomisation of cases works as shown in the user guide. As can be seen, once the user saved a case, the system randomly assigns the case to a court and assigns an ID number (increasing order). Short translation: "For each new case registered, when saved, a number is given and a court is randomly assigned"

A.3 Olympiads Report - Example 1

Matematická olympiáda - obvodné kolo
Kategória MO8 4.4.2017

	meno a priezvisko	adresa školy	1.pr	2pr.	3pr.	body spolu	súťažiaceho pripravoval	
1.	Baksová	*Gymnázium*	6	6	6	18	Mgr. Miklošovičová	úspešný riešiteľ
2.-3.	Kováčová	*ZŠ*	6	3	6	15	Mgr. Kováčová	úspešný riešiteľ
	Šándor	*Gymnázium*	3	6	6	15	Mgr. Miklošovičová	úspešný riešiteľ
4.	Andraščík	*ZŠ*	5	6	2	13	Mgr. Vimpeľová	úspešný riešiteľ
5.-7.	Jurík	*ZŠ s MŠ*	6	6	0	12	Mgr. Pravňanská	úspešný riešiteľ
	Kováčová	*ZŠ*	6	6	0	12	Mgr. Bočevová	úspešný riešiteľ
	Macurová	*Gymnázium*	6	6	0	12	Mgr. Miklošovičová	úspešný riešiteľ
8.	Gladiš	*ZŠ*	5	6	0	11	Mgr. Kováčová	úspešný riešiteľ
9.-10.	Dubovský	*ZŠsMŠ*	4	6	0	10	Ing. Žilinská	úspešný riešiteľ
	Reilly	*ZŠ*	4	6	0	10	Mgr. Vimpeľová	úspešný riešiteľ
11.-12.	Jakuš	*ZŠ*	6	3	0	9	Mgr. Vimpeľová	úspešný riešiteľ
	Fraňo	*Gymnázium*	6	3	0	9	Mgr. Miklošovičová	úspešný riešiteľ
13.-14.	Fraňo	*ZŠ*	5	3	0	8	Mgr. Molnárová	
	Achbergerová	*ZŠ s MŠ*	5	3	0	8	Mgr. Nováková	
15.	Turiničová	*Gymnázium*	6	1	0	7	PaedDr. Vaňková	
16.-17.	Kurčina	*ZŠ*	6	0	0	6	Mgr. Molnárová	
	Agner	*ZŠ s MŠ*	3	3	0	6	Mgr. Nováková	
18.-19.	Gvozdjak	*ZŠ*	5	0	0	5	Mgr. Bočevová	
	Muráriková	*Gymnázium*	2	3	0	5	PaedDr. Vaňková	
20.-22.	Boldiš	*ZŠ s MŠ*	3	1	0	4	Mgr. Pravňanská	
	Gottschallová	*Gymnázium*	4	0	0	4	Mgr. Miklošovičová	
	Lupčová	*Gymnázium*	4	0	0	4	Mgr. Miklošovičová	
23.	Mackovčin	*Gymnázium*	3	0	0	3	Mgr. Miklošovičová	
24.	Šipoš	*ZŠ*	0	0	1	1	Mgr. Molnárová	
25.	Šťastná	*ZŠsMŠ*	0	0	0	0	Ing. Žilinská	

Notes: This figure shows one example of how reports are shown publicly. The results are for the Z7 category (7$^{\text{th}}$ grade) in 2016. The first column shows the rank of participants in the district. The second column shows the full names of participants. The third column shows the school where participants study. Columns 4, 5, 6, and 7 show the points obtained in question 1, 2, 3, and the total score obtained, respectively. Column 8 shows the name of the tutor. The last column shows the label for students who achieved 9 points or more. The translation is "successful participant". All first names of participants and tutors are erased, and names of schools as well.

A.4 Olympiads Report - Example 2

Matematická olympiáda - obvodné kolo
Kategória MO7 5.4.2016

umiestnenie	meno a priezvisko	adresa školy	1.pr	2pr.	3pr.	body spolu	súťažiaceho pripravoval	
1.	Šándor	Gymnázium	6	0	5	11	Mgr. Miklošovičová	úspešný riešiteľ
2.	Kováčová	ZŠ	0	3	6	9	Mgr. Lišková	úspešný riešiteľ
NR 3.	Andraščík	ZŠ	3	4	0	7	Mgr. Vimpeľová	
4.-8.	Boldiš	ZšsMŠ	1	1	4	6	Mgr. Kováčová	
	Fischerová	Gymnázium	3	3	0	6	Mgr. Miklošovičová	
	Mackovčin	Gymnázium	3	3	0	6	Mgr. Miklošovičová	
	Matúš	ZŠ	0	4	2	6	Mgr. Vimpeľová	
	Kurčina	ZŠ	3	3	0	6	Mgr. Molnárová	
9.-10.	Turiničová	Gymnázium	3	0	2	5	PaedDr. Vaňková	
	Jurík	ZšsMŠ	1	2	2	5	Mgr. Kováčová	
11.-14.	Baksová	Gymnázium	3	1	0	4	Mgr. Miklošovičová	
	Fraňo	Gymnázium	0	3	1	4	Mgr. Miklošovičová	
	Tadeáš	ZŠ	1	3	0	4	Mgr. Vimpeľová	
	Batka	ZŠ	0	3	1	4	Mgr. Vimpeľová	
15.	Muráriková	Gymnázium	0	3	0	3	PaedDr. Vaňková	
16.-18.	Balcová	Gymnázium	0	1	1	2	Mgr. Miklošovičová	
	Dubovský	ZšsMŠ	0	1	1	2	Ing. Žilinská	
	Semešová	ZŠ	0	2	0	2	Mgr. Molnárová	
19.-23.	Molinari	Gymnázium	0	1	0	1	PaedDr. Vaňková	
	Kloknerová	Gymnázium	0	1	0	1	PaedDr. Vaňková	
	Krajčovič	Gymnázium	0	0	1	1	PaedDr. Vaňková	
	Gottschalová	Gymnázium	0	1	0	1	Mgr. Miklošovičová	
	Ješko	ZŠ sMŠ	0	1	0	1	Mgr. Nováková	
24.-28.	Belková	Gymnázium	0	0	0	0	PaedDr. Vaňková	
	Krebs	Gymnázium	0	0	0	0	PaedDr. Vaňková	
	Wagnerová	Gymnázium	0	0	0	0	PaedDr. Vaňková	
	Lupčová	Gymnázium	0	0	0	0	Mgr. Miklošovičová	
	Šťastná	ZšsMŠ	0	0	0	0	Ing. Žilinská	

Notes: This figure shows one example of how reports are shown publicly. The results correspond to the Z8 category (8^{th} grade) in 2017. The first column shows the rank of participants in the district. The second column shows the full names of participants. The third column shows the school where participants study. Columns 4, 5, 6, and 7 show the points obtained in question 1, 2, 3, and the total score obtained, respectively. Column 8 shows the name of the tutor. The last column shows the label for students who achieved 9 points or more. The translation is "successful participant". All first names of participants and tutors are erased, and names of schools as well.