INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

# PARENTAL INVOLVEMENT AND EDUCATION OUTCOMES OF THEIR CHILDREN

*Klara Kantova*

IES Working Paper 1/2022

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz



**Bibliographic information**:
Kantova K. (2022): "Parental Involvement and Education Outcomes of Their Children"
IES Working Papers 1/2022. IES FSV. Charles University.

This paper can be downloaded at: http://ies.fsv.cuni.cz

# Parental Involvement and Education Outcomes of Their Children

Klara Kantova

Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic

January 2022

**Abstract:**
With the goal to shed more light on the effectiveness of parental time spent with their children, I estimate the causal relationship between parental involvement and education outcomes of children. My research is the first which examines the effect of parental time in terms of the engagement in child's everyday life. Moreover, I improve the existing literature by including the characteristics of children as well as parents. To estimate causal treatment effects, I use a simple logistic regression along with a subclassification on the propensity score. By subclassification, the systematic differences in baseline characteristics are eliminated. The education outcome is represented by a binary variable denoting whether the respondent completed high school or not. Completing high school improves person's economic selfsufficiency and civic engagement. The results indicate higher probability of graduating from high school with higher parental involvement. Moreover, the probability decreases when child disobeys his or her parent. Most importantly, the results suggest that the expected probability of completing high school decreases with more strict parental behavior.

**JEL:** I21, J12, J13
**Keywords:** Education, Children, Family, Parental Involvement

**Acknowledgements:** This paper is dedicated to project SVV 260 597.

# 1 Introduction

Spending time together is important for family's well-being, especially in the fast pace of modern days. Parental involvement builds self-esteem of children, strengthens family bonds, develops their positive behavior, and encourages communication. Also, it can affect child's academic performance.

Parental time spent with their child is the best investment in extending child's human capital. Mcdowell *et al.* (2018) claim that parental involvement in child's early years has significant effect on their cognitive development, literacy, and number skills. Moreover, Feinstein & Symons (1999) claim that parental involvement in a child's schooling between the ages of 7 and 16 is more powerful than family background, size of a family, or level of parental education in terms of secondary school achievement. Also, child's education failure is increased by lack of parental interest in schooling.

Even though Feinstein & Symons (1999) claim that parental involvement is more powerful than parental education, many studies show that parental education level is correlated with the amount of time spent with children. Studies argue that higher-educated parents actually spend more time with their children. Guryan *et al.* (2008) found that mothers with a college education or higher spend more than 4 hours per week more with their children than mothers with lower education. These results are surprising as the opportunity cost of time is much higher for more educated parents (i.e., higher waged parents according to the author) than for less educated. A study by Guryan *et al.* (2008) also examines whether the observed relationship of parental education and time spent with children in the US holds in other countries too. It examines 13 countries (e.g., Norway, UK, Netherlands, Canada, Chile, South Africa, Palestine, ...) and compares the results with the US. The sample is restricted to the households which includes individuals (adults, mothers, and fathers) aged 21 to 55 inclusive and have at least one child younger than 18 in the household. For almost all observed countries, the same result holds as for the US. Also Kalil *et al.* (2012) tested the hypothesis that highly educated mothers spend more time in an active childcare than mothers with lower education. Their study not only examines that high educated mother spends more time with her child, but also whether the distribution of her time spent in a childcare is more effective than by less educated mother. McLanahan (2004) argues that children of high educated women are gaining assets such as parental time and money. The study was conducted with a purpose that government should deal with increasing inequality between rich and poor and focus on closing this widening gap.

Paper by Rasmussen (2009), which focuses on families with two full-time employed parents, analyzes, theoretically and empirically, relationship between parental time use and child development. The impact of parental involvement on children's education outcome is monitored in Denmark using Danish time use data.

Setting limits with children from the early age helps to develop their discipline. Gained and developed discipline then improve child's future academic performance. According to Morin (2021), there is a difference between setting the limits and rules. The limits express the guideline for behavior and give opportunities to deepen children's skills. If they are established at the early age of a child, it will make later education easier and it becomes a habit. The limits teach children self-discipline, keep them safe and healthy. Moreover, the limits help children to cope with uncomfortable feelings and show them that their parents care. The goal is to teach them to manage all responsibilities, such as homework, chores, brushing teeth, etc., without reminding them to do so.

Even though children often like to test parents how serious they are about the limit, according to Morin (2021), it does not mean they do not want to have those limits. The limits are giving to children the feeling of parental care. Many parents avoid setting the limits as they do not want to make uncomfortable their children as well as themselves. However, with each limit comes the opportunity for the child to try to manage the emotions and to prove that he or she can be responsible with those limits. According to Morin (2020), parents should start teaching their children also the impact of breaking the limits or fulfilling the requests (i.e., punishment or praise) from an early age. They can learn from childhood that good choices as doing chores, homework, etc., lead to positive consequences. Oppositely, bad choices, as is physical aggression, lead to negative consequences.

As stated by Zaff *et al.* (2017), nearly one in five students does not complete high school (i.e., does not graduate) on time, if ever. Completing high school subconsciously teaches the students how to positively contribute to economy as well as to civic life. As claimed by Belfield & Levin (2007), graduation from high school is a doorway to economic self-sufficiency, and civic engagement. Without a high school diploma, people are more likely to earn lower income and to be arrested, which leads to higher costs for the US. According to Sum *et al.* (2009), each high school dropout costs the US about $292,000 more (over lifetime) than high school graduate. It is caused by lower taxable income together with higher reliance on social welfare programs. According to Cole (2017), the more parents are involved in the education of their children, the more students are likely to excel in academic performance. Thus, they become more likely a productive members of the society. Also Bryan (2005) claims that if parents are actively participating in children's education, they are more likely to excel in academics.

Zaff *et al.* (2017) claim that high school dropouts are associated with low-income families, neighborhoods, and ethnicity. Only 72 % of students from low-income families graduate from high school on time (87 % of students from high or middle income families). Concerning the ethnicity, African-American and Hispanic students have about 10 % lower graduation rate than the national average. They also claim that children's graduation is associated with their parents' education. McCallumore & Sparapani (2010) examine the importance of ninth grade on high school graduation. According to their study, the increase of high school dropouts in 2001 occurred when there was a significant emphasis on obtaining college degree. Students felt under the pressure and did not complete high school. The focus of the study is on the ninth grade because the authors claim that the problem of dropouts arises from the transition from middle school to high school. Hoover-Dempsey & Sandler (1997) look at the issue from a different perspective. They studied why parents become involved in their children's education from a psychological point of view.

Most studies dealing with the impact of parental involvement on children's educational outcomes investigate the parental engagement in terms of school - having discussion about the school, helping with homework, reading with children, and concerning also parental involvement at school such as parents' volunteering, attending the workshops, school plays, sport events, being involved in PTA etc. My research contributes to the existing literature in various ways. Firstly, it examines the effect of parental time spent with children on their education level in terms of the engagement of parents in child's everyday life. Secondly, this work contributes to a gap in literature by including the characteristics of children as well as parents. Thirdly, the study assists in better understanding of the role of parents in education outcomes of their children.

The subjects of the research are Americans born between 1983-1984 with both birth parents present in the household when the respondents are 12 to 14 years old. The limitation of the sample is given by the characteristics of the explanatory variables used. Data are collected from the National Longitudinal Survey of Youth 1997 (NLSY97) which is a publicly available database. I provide the empirical results from a simple logistic regression and from subclassification matching method. I regress a binary variable determining whether the respondent completed high school or not on a vector of parental involvement variables. In addition, I include a vector of control variables to control the differences in demographics which may affect not only the probability of graduation from high school but also the variables of parental involvement, i.e., explanatory variables. The level of parental involvement is measured by asking youths and their parents questions relating to monitoring, limit setting, and limit breaking. The results indicate higher probability of graduating from high school with higher parental involvement. Moreover, the probability decreases when child disobeys his or her parent. Most importantly, the results suggest that the expected probability of completing high school decreases with more strict parental behavior meaning that higher limit setting results in a reduction of the probability of graduating from high school.

The remainder of the paper is structured as follows: Section 2 gives an introduction to data used for the analysis. It includes detailed description of used variables and it covers basic demographics and descriptive statistics of the sample. Section 3 describes the used methodology. Section 4 presents results and discussion. Finally, Section 5 concludes the findings.

## 2 Data description

The data comes from NLSY97.[1] NLSY97 is a longitudinal project that follows the lives of a sample of Americans born between 1980-1984. This survey is a part of the National Longitudinal Surveys program. Data observes 8,984 individuals over time. Respondents' ages ranged from 12 to 18 when first interviewed in 1997-1998. Now, there are available 18 rounds of interviews with the last round, round 18, held in 2017-2018, when respondents were 32-38 years old.

The main research question is associated with the relationship between parental involvement and children's education outcomes. The questions about parenting techniques (i.e., parental in-

[1] National Longitudinal Survey of Youth 1997.

volvement) were only asked of those respondents whose age ranged from 12 to 14 years when first interviewed. Only 3,578 respondents fulfill this criterion. However, not all of these respondents answered the questions concerning either education attainment or parental involvement. Eventually, I use a sample of 1,868 observations.

## 2.1 Dependent variable

The NLSY97 data contains information about education outcomes of the respondents.[2] For the purpose of the study, I use the information about the highest degree completed by the respondents as the dependent variable. The variable for the highest degree completed is a cumulative variable. This variable is created for each respondent in a different point of time. It is collected after each respondent finishes his or her studies.

In Table 1 you can see that 1,576 (84.4 %) respondents have received a high school diploma (have completed regular 12 years program). Higher education than that is completed by 804 (43.0 %) individuals. GED[3] is completed by 155 respondents. GED determines whether the respondent has a high school graduate's level of knowledge. However, it is not the equivalent of high school diploma.[4]

Table 1: Highest degree received by children

| | n | % |
|---|---|---|
| None | 137 | 7.3 % |
| GED | 155 | 8.3 % |
| High school diploma | 772 | 41.3 % |
| Associate/Junior college | 173 | 9.3 % |
| Bachelor's degree | 430 | 23.0 % |
| Master's degree | 147 | 7.9 % |
| PhD | 12 | 0.6 % |
| Professional degree | 42 | 2.2 % |
| | 1,868 | 100 %* |

* not equal to exactly 100 % due to rounding.
*Notes*: GED - General Educational Development.

The variable of the respondent's highest degree completed is a categorical variable with 8

[2]The information about their GPA is calculated by the school in its metric for the last year of youth's enrollment. As the metric varies significantly across the US countries, I cannot use this information as the education outcome since those values are incomparable.

[3]General Educational Development.

[4]The US high school diploma is the official credential awarded when a student completes secondary education in the US. Most foreign governments recognize the US high school diploma but may or may not recognize the GED since it is not the official document.

different values. Those 8 values are incomparable. Thus, I introduce a dummy variable whether the respondent completed high school or not. In the US, the most important part of education is the graduation from high school as discussed in the Section 1. In Table 2 you can see the portion of those who have completed high school and those who have not. The sample is unbalanced. Moreover, the portion of males who have completed high school is lower than for females.

Table 2: Graduation from high school

| | n | % |
|---|---|---|
| Completed HS | 1,576 | 84.4 % |
| Did not complete HS | 292 | 15.6 % |
| | 1,868 | 100 % |
| *Males* | | |
| Completed HS | 796 | 81.6 % |
| Did not complete HS | 180 | 18.4 % |
| | 976 | 100 % |
| *Females* | | |
| Completed HS | 780 | 87.4 % |
| Did not complete HS | 112 | 12.6 % |
| | 892 | 100 % |

## 2.2 Explanatory variables

I choose explanatory variables such as monitoring youth by mother as well as by father, limit setting and limit breaking reported by youth as well as by parent.

### 2.2.1 Monitoring

Information about monitoring is obtained by asking the respondents four questions.

- How much does he or she know about your close friends, that is, who they are?
- How much does he or she know about your close friends' parents, that is, who they are?
- How much does he or she know about who you are with when you are not at home?
- How much does he or she know about who your teachers are and what you are doing in school?

Responses to these questions are measured on a 5-point scale.[5] The questions were asked of biological mothers as well as fathers. After the responses are collected, the monitoring variable

[5] 0 - parent knows nothing, 1 - knows just a little, 2 - knows some things, 3 - knows most things, 4 - knows everything.

is created by summing these responses. Thus, the value ranges from 0 to 16 where higher score indicates greater parental monitoring.

Buchanan *et al.* (1992) claim that degree of monitoring has impact on child's scholastic achievement, behavior, or even on sexual involvement. In other words, lower degree of parental monitoring is linked with lower education, behavioral problems, and early sexual involvement.

### 2.2.2 Autonomy, Control, and Limit setting

Information about autonomy, control, and limit setting is obtained by asking the respondents three questions. Parallel questions were asked of the parents.

- Who set the limits on how late you stay out at night? / how late he or she can stay out at night?
- Who set the limits on who you can hang out with? / who he or she can hang out with?
- Who set the limits on what kinds of TV shows or movies you can watch? / what kinds of TV shows and movies he or she can watch?

Responses to these questions are measured on a 3-point scale.[6] Then the limit setting variables are created by summing these responses. Values of the limit setting variables (responded by youth as well as parent) range from 0 to 6 where higher score indicates greater parental role in the limit setting. The value of 0 indicates that youth sets all limits and 6 that parent sets all limits.

According to Erford (1995), observing parallel questions is very useful as discrepancies across the answers indicate misunderstanding about who in fact sets the limits, which often leads to limits breaking by youths from the parents' point of view. However, this study does not treat uniquely those discrepancies.

### 2.2.3 Limit breaking

Information about the limit breaking is obtained by asking the respondents three questions. Parallel questions were asked of the parents.

- In the past 30 days, how many times have you broken the limits about how late you can stay out at night? / how many times do you think he or she has broken the rules about how late he or she can stay out at night?
- In the past 30 days, how many times have you broken the limits about who you can hang out with? / how many times do you think he or she has broken the rules about who he or she can hang out with?

[6] 0 - parent let youth decide, 1 - parents and youth decide jointly, 2 - parent or parents set limits.

- In the past 30 days, how many times have you broken the limits about what kinds of TV shows and movies you watch? / how many times do you think he or she has broken the rules about what kinds of TV shows and movies he or she can watch?

Responses to these questions are measured on a 3-point scale.[7] Again, the discrepancies are observed here. About 13 % of youths claim that they did break at least one of the limits even though their parent reported that they did not break any of these three limits.

### 2.2.4 Summary

Table 3 summarizes the name, type and usage of chosen variables. Dependent variable represents whether the respondent received a high school diploma and explanatory variables are designing the level of parental involvement.

Table 3: Description and summary statistics of the regression variables

| Variable | Description | Mean | SD | Median |
|---|---|---|---|---|
| *Dependent variable* | | | | |
| HS | Binary variable whether the respondent received a high school diploma (have completed regular 12 years program) or not. | See Table 2. | | |
| *Explanatory variables* | | | | |
| $Mon_M$ | How much does mother monitor her kid. What does she know about kid's friends, his/her parents, teachers and where her kid is. Min = 1, Max = 16. | 10.75 | 3.08 | 11.00 |
| $Mon_F$ | How much does father monitor his kid. What does he know about kid's friends, his/her parents, teachers and where his kid is. Min = 0, Max = 16. | 8.51 | 3.99 | 9.00 |
| $Lim_Y$ | Variable describing autonomy, control & limit setting. Responded by youth. Min = 1, Max = 6. | 3.42 | 1.44 | 3.00 |
| $Lim_P$ | Variable describing autonomy, control & limit setting. Responded by parent. Min = 1, Max = 6. | 4.33 | 1.25 | 4.00 |
| $Broke_Y$ | How many times have youth broken the limit in past 30 days. Responded by youth. Min = 0, Max = 1. | 0.43 | 30.50 | 0.00 |
| $Broke_P$ | How many times have youth broken the limit in past 30 days. Responded by parent. Min = 0, Max = 1. | 0.30 | 0.46 | 0.00 |

*Notes*: SD = standard deviation.

### 2.2.5 Modification of the variables

The variables concerning parenting techniques are designing the level of parental involvement in child's life. Hence, I treat them as factors. For the purpose of the analysis, I create a dummy variable equal to 1 with high parental involvement, and equal to 0 otherwise. I set the boundaries

[7] 0 - youth did not break any of three limits, 1 - youth broke any of three limits, or 9 - youth sets all three limits, meaning that there are no limits.

based on the definitions of parental involvement variables. Concerning the monitoring variables (by mother, $Mon_M$, and father, $Mon_F$), the responses with values either 0, 1, or 2 (parent knows nothing, knows just a little, knows some things) represent low parental involvement, and the responses with values 3 and 4 (parent knows most things, parent knows everything) represent high parental involvement. Regarding the limit setting variables reported by youths ($Lim_Y$) and parents ($Lim_P$), low level of parental involvement is represented by the values from 0 to 3. Higher values than 3 represent high level of parental involvement. The variables showing if youth have broken any limit in the past 30 days ($Broke_Y$ and $Broke_P$) already were dummy variables. They are equal to 1 if children have broken any limit, and 0 otherwise. The value of 9, meaning that youth sets all three limits, has not been reached by a single respondent in the sample as you can see in Table 3. In Table 9 you can see the number of respondents in control (low parental involvement) and treated (high parental involvement) group for each of the parental involvement variable.

### 2.3 Control variables

Even though Feinstein & Symons (1999) argue that parental involvement is more powerful in terms of succeeding in a secondary school than family background, size of a family, or level of parental education, there is number of studies mentioned in Section 1 which show that parental education, neighborhood or household size affect child's education outcome. Based on the literature, I introduce a vector of control variables. The vector contains variables for sex, ethnicity, location (i.e., neighborhood), and education of both parents. Concerning the sample, percentages of the highest grade completed by biological parents of the respondents are reported in Table 4.

Table 4: Highest degree completed by parents (%)

| | Mothers | Fathers | | Mothers | Fathers |
|---|---|---|---|---|---|
| $2^{nd}$ grade | 0.32 | 0.16 | $1^{st}$ year college | 8.40 | 6.37 |
| $3^{rd}$ grade | 0.54 | 0.64 | $2^{nd}$ year college | 13.87 | 10.55 |
| $4^{th}$ grade | 0.54 | 1.02 | $3^{rd}$ year college | 8.41 | 2.52 |
| $5^{th}$ grade | 0.75 | 0.91 | $4^{th}$ year college | 13.60 | 13.17 |
| $6^{th}$ grade | 2.30 | 2.68 | $5^{th}$ year college | 3.21 | 1.98 |
| $7^{th}$ grade | 0.59 | 1.02 | $6^{th}$ year college | 4.12 | 4.60 |
| $8^{th}$ grade | 2.09 | 2.36 | $7^{th}$ year college | 0.64 | 1.45 |
| $9^{th}$ grade | 3.00 | 3.32 | $8^{th}$ year college or more | 1.23 | 2.94 |
| $10^{th}$ grade | 4.01 | 3.53 | | | |

Continued on next page

Table 4: Categorical control variables (%) (continued)

| | Mothers | Fathers | | Mothers | Fathers |
|---|---|---|---|---|---|
| $11^{th}$ grade | 3.91 | 4.28 | | | |
| $12^{th}$ grade | 34.48 | 36.51 | | | |

From Table 4 you can see that among both fathers and mothers the most often represented level of education is $12^{th}$ grade which corresponds to completing high school. 88.0% of mothers and 80.1% of fathers graduated from high school and either finished their studies or continued.

In Table 5 I denote the number of respondents in a particular group and percentage of the total for categorical control variables. In Table 6 you can observe description and summary statistics of the quantitative control variables within the sample.

Table 5: Categorical control variables

| | n | % |
|---|---|---|
| *Sex* | | |
| Male | 976 | 52.2 |
| Female | 892 | 47.8 |
| | 1,868 | 100 |
| *Ethnicity* | | |
| Black | 297 | 16.0 |
| Hispanic | 359 | 19.2 |
| Mixed race (Non-Hispanic) | 10 | 0.5 |
| Non-Black/Non-Hispanic | 1202 | 64.3 |
| | 1,868 | 100 |
| *Location* | | |
| Urban | 1,290 | 69.0 |
| Rural | 498 | 26.7 |
| Unknown | 80 | 4.3 |
| | 1,868 | 100 |

Table 6: Description and summary statistics of the control variables

| Variable | Description | Mean | SD | Median |
|---|---|---|---|---|
| HH size | The number of participants in a household where respondent grew. Min = 2, Max = 14. | 4.72 | 1.30 | 4.00 |
| hgc M | Mother's highest level of education. Min = 2, Max = 20. | 12.89 | 2.98 | 12.00 |
| hgc F | Father's highest level of education. Min = 2, Max = 20. | 12.85 | 3.27 | 12.00 |

*Notes*: HH - household; hgc M - highest grade completed by mother; hgc F - highest grade completed by father.

# 3 Methodology

## 3.1 Model specification

The goal of the analysis is to estimate the effect of parental involvement on children's education outcome. The following equation presents the baseline model:

$$y_i = \beta_0 + ParentalInvolvement_i'\gamma + X_i'\delta + u_i, \tag{1}$$

where $y_i$ is the dependent variable corresponding to an indicator that respondent $i$ has completed high school, and $i = 1, ..., $ N, where N is the number of respondents. $ParentalInvolvement_i$ is a vector of variables measuring parental involvement of respondent $i$, $X_i$ is a vector of control variables for respondent $i$, i.e., other individual or their parents' characteristics that might influence education outcome, and $u_i$ is the unobserved error.

The outcome variable, child's education outcome, is a binary variable determining whether the respondent completed high school or not. To model dichotomous dependent variable, I choose logit model. The endogeneity problem may occur since the variables measuring parental involvement might be correlated with the high school graduation variable as well as with the error term. Endogeneity causes the coefficients to be biased. Therefore, I include a vector of control variables. I choose control variables based on a study by Bogenschneider (1997). The author provides a clear summary of the literature dealing with the characteristics which affect child's education attainment. Those characteristics are sex, parent's education, family structure, and ethnicity. We already control the family structure as we are dealing only with the families where both biological parents are present. Moreover, Zaff *et al.* (2017) associate high school dropouts with neighborhoods and ethnicity. Thus, our control vector contains sex, ethnicity, size of the household, location (urban/rural), and parental level of education.

To deal with potential nonlinear relationship between control variables and the outcome variable and to be sure that I compare otherwise similar children whose parents are either highly or weakly involved, I introduce a matching method by matching similar respondents according to their characteristics along with a simple logistic regression.

## 3.2 Matching Approach

In matching method terminology, I am interested in the causal effect of the treatment (T=1), "high parental involvement", relative to no treatment (T=0), "low parental involvement", on the child's education outcome.[8] For parental involvement I use the vector of explanatory variables, i.e., $ParentalInvolvement_i$. There is no possibility to observe outcome for a single respondent with both treatment and no treatment as single respondent is either treated (parents are very involved) or not treated (parents are not very involved). That is why I have to equate as much as possible the distribution of covariates among treated and control groups.[9] Rosenbaum & Rubin (1983) called this estimation of the causal effect as a missing data problem, since always one value for each individual is missing. This unobserved outcome is called counterfactual outcome.

### 3.2.1 Potential outcome approach

As discussed in Caliendo & Kopeinig (2005), the main focus should be on individuals and their characteristics – vector of covariates (characteristics of respondents as well as their parents), treatment (level of parental involvement), and outcome (completed high school or not). Respondents with high parental involvement are indicated as treated (T=1), control (T=0) otherwise. The outcomes are then $Y_i(T_i)$ for each individual $i$, where $i$ = 1, ..., N and N is number of respondents. Then the treatment effect can be understood as:

$$\tau_i = Y_i(1) - Y_i(0) \tag{2}$$

In Equation 2, $\tau_i$ represents individual causal effect, the treatment effect for an individual $i$. As I discussed before, for each individual $i$ only one of the potential outcomes can be observed. Here occurs the fundamental evaluation problem. Since I cannot observe both outcomes for each individual, I am interested in average treatment effects (or population average treatment effects). To estimate average treatment effects, the treatment effect for individual $i$ has to be independent of others' treatment participation.

[8]The estimation is similar to analyzing treatment effect in medicine. The treatment variable is represented by parental involvement variables.

[9]Vector of covariates is equal to vector of control variables.

### 3.2.2 Parameter of interest

The most common primary treatment effects of interest are ATT, which is defined as the *Average Effect of the Treatment* for those who receive the treatment, and ATE, which is defined as the *Average Treatment Effect in the Population.* They are called estimands. The estimands control how the subclasses are created and how the weights are computed. I use ATT estimand, i.e., subclassification is based on quantiles of the distance measure in the treated group. ATT is defined as:

$$\tau_{ATT} = E(\tau|T=1) = E[Y(1)|T=1] - E[Y(0)|T=1] \tag{3}$$

The problem arises here. $E[Y(0)|T=1]$ represents the unobserved - counterfactual - mean for those being treated. To complete an estimation of ATT, I have to substitute the counterfactual mean for those being treated. Individuals are selected into treatment groups by many factors that may or may not influence the outcome. That is why using $E[Y(0)|T=0]$ (mean outcome of untreated) is not appropriate as the factors are most likely to affect the treatment decision as well as the outcome variable I am interested in. In other words, the outcomes from both groups would differ even if I did not introduce the treatment variable, which would lead to a self-selection bias. By regrouping Equation 3 and adding $-E[Y(0)|T=0]$ to both sides of the equation I obtain:

$$E[Y(1)|T=1] - E[Y(0)|T=0] = \tau_{ATT} + \underbrace{E[Y(0)|T=1] - E[Y(0)|T=0]}_{\text{self-selection bias}}, \tag{4}$$

where parameter $\tau_{ATT}$ is defined only if $E[Y(0)|T=1] - E[Y(0)|T=0] = 0$, meaning that the self-selection bias does not occur. To solve the problem in Equation 4, I introduce propensity score matching.

### 3.2.3 Propensity score

The propensity score ($P(\mathbf{X}) = P(T=1|\mathbf{X})$) is the probability of being treated for an individual given the covariates $\mathbf{X}$. The propensity score is one of the balancing scores. According to Rosenbaum & Rubin (1983), if potential outcomes are independent of the treatment conditional on the covariates $\mathbf{X}$, then they are also independent of the treatment conditional on a b($\mathbf{X}$), the balancing score.

Assumptions for estimating strategy:

1. **Positivity (Common Support):** assignment is probabilistic:

$$0 < P[T_i = 1|\mathbf{X}, Y(1), Y(0)] < 1,$$

or

$$0 < P[T_i = 1|\mathbf{X}] < 1.$$

2. **No unmeasured confounding:**

$$P[T_i|\mathbf{X}, Y(1), Y(0)] = P[T_i|\mathbf{X}],$$

further - unconfoundedness given propensity score:

$$P[T_i|P(\mathbf{X})].$$

First assumption is also known as overlap condition. Let $\mathbf{X}$ be a set of observable covariates which are not affected by the treatment (i.e., control variables). Common support ensures that individuals with the same covariates' values have positive probability of being both - treated and non treated. If P($\mathbf{X}$)=0 or P($\mathbf{X}$)=1 for some $\mathbf{X}$, then the individuals with such $\mathbf{X}$ are either never treated or always treated and I cannot use matching. Second assumption, also called conditional independence assumption or unconfoundedness, assumes that potential outcomes are independent of treatment conditional on vector of covariates $\mathbf{X}$, or further on propensity score P($\mathbf{X}$).

If both assumptions hold, the propensity score matching estimator for ATT is simply the mean difference of outcomes weighted by respective propensity scores. When estimating ATT, sufficient assumptions are $P[T_i = 1|\mathbf{X}] < 1$, and $P[T_i|\mathbf{X}, Y(0)] = P[T_i|P(\mathbf{X})]$. $\tau_{ATT}^{PSM}$ is then mathematically defined as:[10]

[10]PSM - Propensity Score Matching.

$$\tau_{ATT}^{PSM} = E_{P(\mathbf{X})|T=1}\{E[Y(1)|T=1, P(\mathbf{X})] - E[Y(0)|T=0, P(\mathbf{X})]\}. \tag{5}$$

**Estimating the Propensity score.** According to Caliendo & Kopeinig (2005), the model of propensity score should include only variables which are not affected by the treatment, and the covariates $\mathbf{X}$ have to satisfy assumption (2). Also, omitting important variables leads to a bias in the resulting estimates. In other words, I include variables which affect both the treatment decision and the outcome variable. The best way how to prevent violating these assumptions is to choose variables which are fixed over the time or which are measured before the treatment. Chosen set of covariates meets these conditions, since I do not assume that sex, ethnicity, and the highest degree completed by parents are likely to change. Concerning household size and location, those covariates could change. However, they are measured at the same point of time as the treatment variables. With regard to model choice, I use logit model for estimating the propensity score.

### 3.2.4 Subclassification

I perform the propensity score subclassification using the `MatchIt` package in R.[11] Unlike other matching methods, subclassification uses all individuals from the treatment as well as from the control group. Binary outcomes are measured by the risk difference, risk ratio, or odds ratio (OR). The OR is a noncollapsible effect measure, so the computation of marginal effect estimate is done by computing the average of the predicted subclass-specific odds under each treatment from which is then computed the marginal effect estimate.

Orihara & Hamada (2021) discuss the optimal number of subclasses for subclassification on the propensity score. They select the number of subclasses which minimizes MSE of the subclassification estimator.[12] According to Rosenbaum & Rubin (1983), 90% of bias is removed by only five subclasses based on the propensity score. Five subclasses are not universal recommendation and may not always be optimal. However, in this case, it is suitable. I choose five subclasses based on the balance assessment (matches diagnostics).

[11]For a more detailed description of MatchIt package in R see: `https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf`

[12]MSE - Mean Squared Error.

**Balance Assessment.** Diagnosing the quality of matched samples is the most important step when using the matching method. When matching is done, it should be followed by evaluation of the covariate balance whether the groups are matched correctly. I use standardized mean differences (SMD) for balance assessment proposed by Austin *et al.* (2005) and Austin *et al.* (2007). The SMD is the difference in means of each covariate between treated and control groups standardized by a standardization factor. When targeting ATT, the standardization factor is the standard deviation of the covariate in the treated group. Mathematically:

$$d = \frac{100 \times |\overline{x}_T - \overline{x}_C|}{\sqrt{\frac{s_T^2 + s_C^2}{2}}}, \tag{6}$$

where $d$ is the standardized difference. $\overline{x}_T$ and $\overline{x}_C$ are the means of the variables among the treated and control subjects, respectively. $s_T^2$ and $s_C^2$ are the sample standard deviantions of covariates in the treated and control subjects, respectively.

In Table 7 you can see the standardized mean difference for each covariate prior to matching and for matched sample. The value of standardized mean difference close to zero indicates successful balance. Vignettes for `MatchIt` package recommend thresholds 0.1 and 0.05 for prognostically important covariates. As the SMDs of all covariates after the matching are below 0.1, even below 0.05 except the SMD of highest grade completed by mother for limit setting variable responded by parent, the balance of the means between treated and control group significantly improved after the matching. Furthermore, I present graphical diagnostics (Figure 1 - love plot of SMD, Figure 2 - distribution of propensity scores, and Figure 3 - histograms of propensity scores) for the variable *monitoring by father*. You can find graphical diagnostics of all parental involvement variables in the Appendix.

**Robust and Cluster-Robust Standard Errors.** Usage of the cluster-robust standard errors is preferred in the most cases of the matching method. They are appropriate when a large number of clusters is present. With subclassification method, there is no large number of clusters. In the case of subclassification method, regular robust standard errors are appropriate when estimating marginal effects.[13]

[13]For more details about estimating effects after matching and (cluster-) robust standard errors see vignettes for MatchIt package: `https://cran.r-project.org/web/packages/MatchIt/vignettes/estimating-effects.html`

Table 7: Standardized mean difference

| | $Mon_M$ | $Mon_F$ | $Lim_Y$ | $Lim_P$ | $Broke_Y$ | $Broke_P$ |
|---|---|---|---|---|---|---|
| SMD BEFORE SUBCLASSIFICATION | | | | | | |
| distance | 0.4601 | 0.3605 | 0.3208 | 0.3107 | 0.2523 | 0.3884 |
| sexM | -0.1369 | 0.0631 | 0.0835 | 0.0565 | 0.1794 | 0.3575 |
| sexF | 0.1369 | -0.0631 | -0.0835 | -0.0565 | -0.1794 | -0.3575 |
| raceblack | -0.2354 | -0.2168 | 0.1327 | 0.2349 | 0.0229 | 0.0952 |
| racehispanic | -0.1352 | -0.0996 | 0.0743 | 0.0590 | 0.0953 | -0.0470 |
| racemixed | -0.0370 | -0.0059 | -0.0617 | 0.0165 | -0.0108 | -0.0420 |
| racenonBH | 0.2911 | 0.2387 | -0.1598 | -0.2367 | -0.0963 | -0.0340 |
| HH size | -0.0940 | -0.0302 | 0.1414 | 0.0934 | 0.0970 | 0.0239 |
| urban0 | 0.1105 | 0.1820 | 0.0668 | 0.0182 | -0.0937 | -0.0806 |
| urban1 | -0.0624 | -0.1647 | -0.0849 | -0.0262 | 0.0830 | 0.0814 |
| urban2 | -0.1066 | -0.0282 | 0.0459 | 0.0199 | 0.0130 | -0.0110 |
| hgc M | 0.2609 | 0.1370 | -0.2067 | -0.1215 | -0.1058 | 0.0181 |
| hgc F | 0.2696 | 0.2266 | -0.2059 | -0.1348 | -0.0988 | -0.0273 |
| SMD AFTER SUBCLASSIFICATION | | | | | | |
| distance | 0.0426 | 0.0168 | 0.0163 | 0.0155 | 0.0279 | 0.0312 |
| sexM | -0.0191 | 0.0000 | -0.0009 | -0.0078 | 0.0057 | 0.0210 |
| sexF | 0.0191 | -0.0000 | 0.0009 | 0.0078 | -0.0057 | -0.0210 |
| raceblack | -0.0168 | -0.0274 | 0.0004 | 0.0440 | 0.0049 | 0.0198 |
| racehispanic | -0.0128 | 0.0016 | 0.0157 | -0.0400 | 0.0232 | -0.0009 |
| racemixed | 0.0093 | 0.0074 | -0.0160 | -0.0140 | -0.0075 | -0.0114 |
| racenonBH | 0.0216 | 0.0172 | -0.0115 | 0.0001 | -0.0222 | -0.0138 |
| HH size | 0.0209 | 0.0046 | 0.0170 | -0.0055 | 0.0267 | -0.0023 |
| urban0 | -0.0006 | -0.0196 | -0.0020 | 0.0091 | -0.0028 | -0.0286 |
| urban1 | -0.0056 | 0.0171 | 0.0005 | 0.0044 | 0.0043 | 0.0272 |
| urban2 | 0.0149 | 0.0046 | 0.0030 | -0.0297 | -0.0036 | 0.0000 |
| hgc M | 0.0322 | 0.0002 | -0.0178 | 0.0506 | -0.0244 | 0.0038 |
| hgc F | 0.0207 | 0.0208 | -0.0049 | 0.0248 | -0.0258 | 0.0092 |

*Notes:* Horizontal: $Mon_M$ - monitoring by mother; $Mon_F$ - monitoring by father; $Lim_Y$ - limit setting reported by youth; $Lim_P$ - limit setting reported by parent; $Broke_Y$ - limit breaking reported by youth; $Broke_P$ - limit breaking reported by parent. Vertical: sexM - male; sexF - female; racenonBH - Non-Black/Non-Hispanic; HH size - household size; urban0 - rural; urban1 - urban; urban2 - unknown; hgc M - highest grade completed by mother; hgc F - highest grade completed by father.

Figure 1: Love plot - Monitoring by father

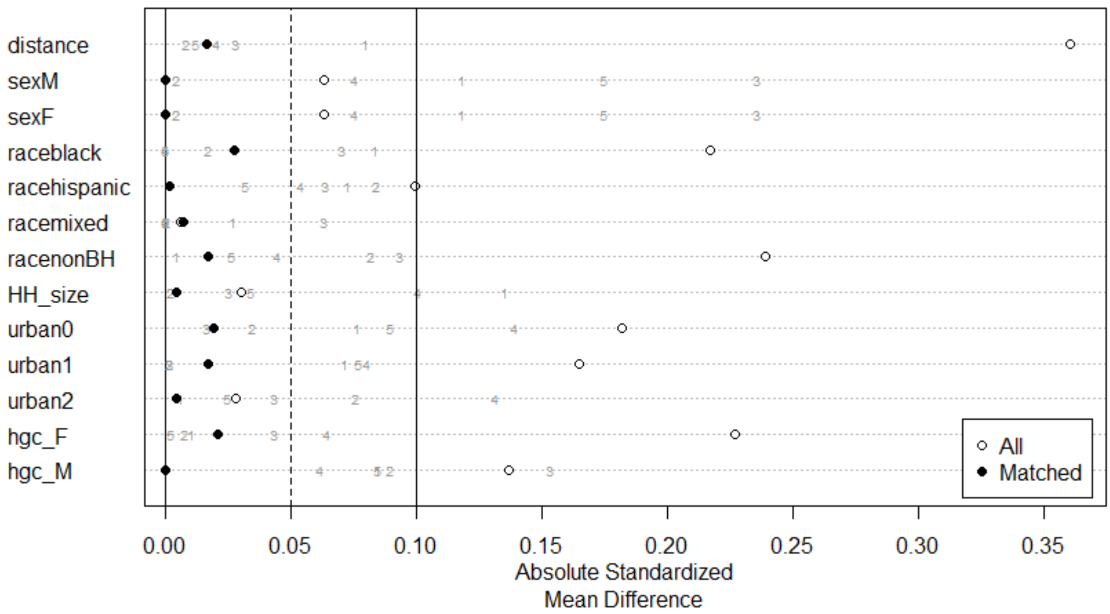



Figure 2: Monitoring by father

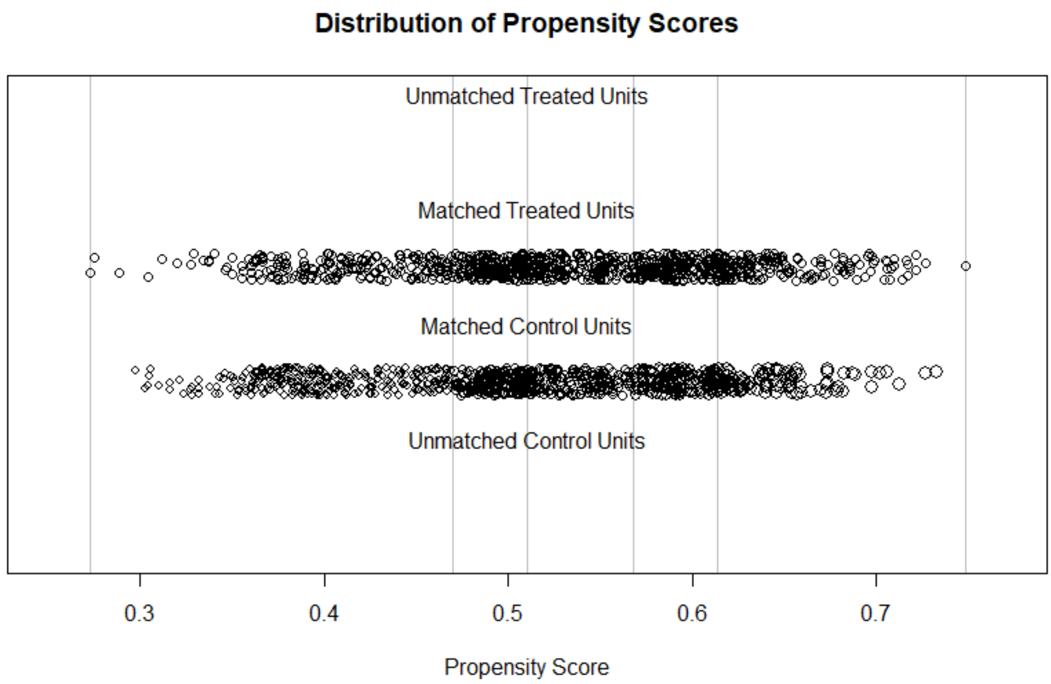
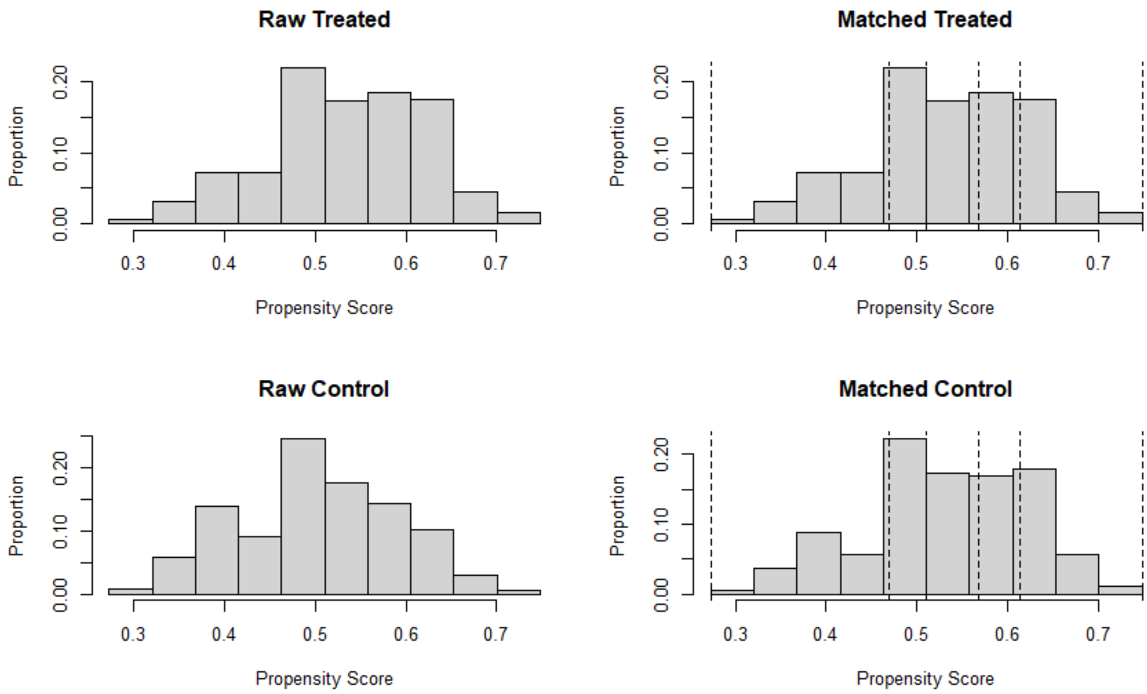


Figure 3: Propensity Scores Histograms - Monitoring by father




# 4 Results

## 4.1 Sample description

The sample consists of 1,868 participants. From a simple data observing you can see that mothers monitor their children more than fathers. The limit setting variable reported by respondents is about 3 % higher than the same variable reported by parents. Also, more respondents reported limit breaking than their parents.

Regarding the sample, females are more likely to graduate from high school. About 10.7 % of Black women and nearly 42 % of Black men did not complete high school. Concerning Hispanic race, the portions are 20.5 % and 33.6 % for women and men, respectively. Nearly 13.8 % of White women and 16.5 % of White men did not graduate from high school. Speaking of mixed race, the portions are not good indicators as there are only 10 respondents of a mixed race (100 % of women and 50 % of men of mixed race did complete high school).

The data suggests that the indicator for completing high school tends to differ with respect to respondents' observable characteristics. As other have shown, respondents who graduated from high school tend to have more educated parents and are more likely living in a household with less persons. In Table 8 you can see means with SDs in parentheses for respondents who have or have not finished high school with respect to household size and highest grade completed by parents. The means of parental education are higher for those respondents who have completed high school.

Table 8: Mean and SD by high school status

| | HH size | hgc M | hgc F |
|---|---|---|---|
| **High school**: | | | |
| yes | 4.69 | 13.17 | 13.17 |
| | (1.26) | (2.91) | (3.20) |
| no | 4.90 | 11.40 | 11.11 |
| | (1.48) | (2.91) | (3.12) |

*Notes:* HH size - household size (range: <2;14>); hgc M - highest degree completed by mother (range: <2;20>); hgc F - highest degree completed by father (range: <2;20>).

In addition, I provide Table A2 in Appendix. It depicts the comparison of means and standard deviations of quantitative control variables for those who are treated (high parental involvement = 1) and those who are not (low parental involvement = 0), separately for males and females. Concerning the variable of monitoring, more educated parents tend to monitor their children more.[14] Moreover, the results show that less educated parents are setting more limits. Respondents who broke the limit tend to have less educated parents. In addition, less educated parents think that their child did not break the limit more than educated parents. Speaking of the household size, there is no observable pattern.

## 4.2 Simple Logistic Regression

Simple logistic regression estimates the effects of parenting styles, keeping other child and parental characteristics constant, without caring for the common support. The logistic regressions are performed separately for each parental involvement variable. The dependent variable is a binary education variable determining whether the respondent graduated from high school or not. I choose as a threshold for statistical significance a p-value less than 0.05 (i.e., 5 % level of significance).

[14]Different result is observed only for the highest degree completed by mother when we look at father's monitoring variable for female respondents.

**Odds ratios.** The odds ratios of parental involvement variables and their confidence intervals are reported in Table 11.[15] The CI for the variable of limit setting reported by youths includes the value of 1 meaning that the result is not statistically significant. The CIs for monitoring youth by father and limit setting reported by parents are close to the value of 1. The values of the exponentiated coefficients for parental monitoring are greater than 1, meaning that the odds of graduating from high school increase when respondent is monitored by mother or father. Also, the effect is greater when youth is monitored by mother than by father. There is a 38-61% increase in the odds of graduation from high school when monitored by parents. Concerning the limit breaking variables, there is a 39-42% decrease in the odds of graduating from high school when respondent breaks the limit. Setting the limits to the respondents decreases the odds of graduation from high school. However, only the limit setting variable responded by parents is statistically significant.

**Marginal Effects.** Marginal effects are depicted also in Table 11. The base level refers to "low parental involvement". The marginal effect of statistically significant coefficients ranges from -0.07 to 0.06. The results show that the expected probability of completing high school is higher for those who are monitored by parents (i.e., treated) than for those who are not monitored. The probabilities are higher by 6 and 4 percentage points when monitored by mother and father, respectively. Limit breaking decreases the probability of completing high school by 6 to 7 percentage points.

### 4.3 Subclassification

**Estimating the Propensity Score.** As I described in Section 3, the systematic differences in baseline characteristics between respondents in the treated and the control group are reduced or eliminated by matching. For the estimation of the propensity score I use logit model and five subclasses. In Table 9 you can see the number of respondents in each group for each treatment variable. Further, in Table 10 you can observe the number of respondents in each subclass and respective subclasses' weights.

[15]The vector of control variables was included in the regression, however, I present only the coefficients concerning the explanatory variables. The estimates of control variables are depicted in Table A3 in Appendix.

Table 9: Number of control vs treated

| | $Mon_M$ | $Mon_F$ | $Lim_Y$ | $Lim_P$ | $Broke_Y$ | $Broke_P$ |
|---|---|---|---|---|---|---|
| Control | 408 | 897 | 994 | 453 | 1056 | 1311 |
| Treated | 1460 | 971 | 874 | 1415 | 812 | 557 |

*Notes:* $Mon_M$ = monitoring by mother; $Mon_F$ = monitoring by father; $Lim_Y$ = limit setting, youth report; $Lim_P$ = limit setting, parent report; $Broke_Y$ = limit breaking, youth report; $Broke_P$ = limit breaking, parent report.

Table 10: Weights & number of respondents

| $Mon_M$ | | $Mon_F$ | | $Lim_Y$ | | $Lim_P$ | | $Broke_Y$ | | $Broke_P$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Weight | N | Weight | N | Weight | N | Weight | N | Weight | N | Weight | N |
| 0.565 | 144 | 0.635 | 281 | 0.620 | 321 | 0.702 | 129 | 0.721 | 285 | 0.666 | 396 |
| 0.995 | 82 | 0.919 | 196 | 0.994 | 199 | 0.749 | 121 | 0.888 | 243 | 0.734 | 356 |
| 1 | 1460 | 0.979 | 184 | 1 | 874 | 1 | 1415 | 1 | 812 | 1 | 557 |
| 1.063 | 77 | 1 | 971 | 1.048 | 190 | 1.147 | 79 | 1.009 | 210 | 1.215 | 215 |
| 1.329 | 61 | 1.210 | 139 | 1.214 | 164 | 1.224 | 74 | 1.262 | 167 | 1.233 | 208 |
| 1.867 | 44 | 1.962 | 97 | 1.659 | 120 | 1.812 | 50 | 1.404 | 151 | 1.973 | 136 |

*Notes:* $Mon_M$ = monitoring by mother; $Mon_F$ = monitoring by father; $Lim_Y$ = limit setting, youth report; $Lim_P$ = limit setting, parent report; $Broke_Y$ = limit breaking, youth report; $Broke_P$ = limit breaking, parent report.

Initial propensity score models are estimated by using the vector of covariates. The estimated propensity scores are predicted probabilities of exposure to the treatment (parental involvement) from the logistic regression models. Once the propensity scores have been estimated for each subject, treated and control subjects are matched on the respective propensity scores. This is performed independently for each variable of parental involvement, i.e., I obtain six propensity score models.

**Odds ratios.** In Table 11 are reported the odds ratios and the CIs after the subclassification. The CIs for the variables of limit setting reported by both youths and parents include the value of 1 - the results are not statistically significant. Speaking of the variable of monitoring by father, the lower bound is close to 1. There is a 68% increase in the odds of graduating from high school when monitored by mother. As for fathers, the increase in the odds is by 32%. Speaking of the limit breaking variables, there is a 39-41% decrease in the odds of graduation from high school when respondent breaks the limit.

**Marginal Effects.** In Table 11 you can also see the marginal treatment effects with the robust standard errors in parentheses after the subclassfication. The base level refers to "low parental involvement". All of the observed coefficients are statistically significant and they range from -0.08 to 0.10. The treatment represented by monitoring variables increases the expected probability of completing high school. The variable of monitoring youth by mother shows the largest marginal effect of 0.10. Meaning that the expected probability of youth completing high school is greater by 10 percentage points for those who are monitored by mother. When monitored by father, the expected probability of youth completing high school is greater only by 6 percentage points. Breaking the limits decreases the expected probability of youth completing high school by 8 percentage points for both observed variables.

## 4.4 Comparison

In Table A3 and Table A4 in Appendix you can find model statistics depicting fit of the simple logistic regression and subclassification regression, respectively. When comparing results from the simple logistic regression and from the regression with matching, the effects are greater with matching. Moreover, the results are more statistically significant when estimated by matching method. The surprising signs of the marginal effects of the limit setting variables hold in both cases. However, when reported by youth, the significant effect is observed only by matching.

Table 11: Results

| **Dependent variable**: | LOGIT | | | | MATCHING | | | |
|---|---|---|---|---|---|---|---|---|
| HS | OR | CI | | ME | OR | CI | | ME |
| | | 2.5% | 97.5% | (SE) | | 2.5% | 97.5% | (SE) |
| **Explanatory variables**: | | | | | | | | |
| Parental involvement | | | | | | | | |
| $Mon_M$ | 1.61 | 1.20 | 2.15 | $0.06^{**}$ (0.02) | 1.68 | 1.26 | 2.23 | $0.10^{***}$ (0.02) |
| $Mon_F$ | 1.38 | 1.06 | 1.80 | $0.04^{*}$ (0.02) | 1.32 | 1.02 | 1.71 | $0.06^{**}$ (0.02) |
| $Lim_Y$ | 0.89 | 0.69 | 1.16 | -0,01 (0.02) | 0.91 | 0.71 | 1.16 | $-0.04^{*}$ (0.02) |
| $Lim_P$ | 0.68 | 0.48 | 0.95 | $-0.04^{*}$ (0.02) | 0.74 | 0.54 | 1.00 | $-0.05^{**}$ (0.02) |
| $Broke_Y$ | 0.61 | 0.47 | 0.79 | $-0.06^{***}$ (0.02) | 0.61 | 0.48 | 0.78 | $-0.08^{***}$ (0.02) |
| $Broke_P$ | 0.58 | 0.44 | 0.76 | $-0.07^{***}$ (0.02) | 0.59 | 0.46 | 0.77 | $-0.08^{***}$ (0.02) |

*Notes:* $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Coefficient (marginal effect dy/dx) for categorical variable is the discrete change from the base. Robust standard errors are provided in parentheses. OR - odds ratio, CI - confidence interval, ME - marginal effect, SE - standard error. $Mon_M$ - monitoring by mother; $Mon_F$ - monitoring by father; $Lim_Y$ - limit setting, youth report; $Lim_P$ - limit setting, parent report; $Broke_Y$ - limit breaking, youth report; $Broke_P$ - limit breaking, parent report.

### 4.5 Discussion

The problem in the estimation of the effect of parental involvement on the child's education outcome is that there might be an issue that more educated parents use more appropriate parenting techniques. They may set the limits more accordingly to the child. Thus, children of more educated parent might have less intentions to break the limits.

The effect of parental education on child's behavior can be observed on innate (e.g., IQ) as well as on acquired characteristics. Guryan *et al.* (2008) and Kalil *et al.* (2012) suggest that more educated parents spend more time with their children than less educated parents. Moreover, the study by Neidell (2000) claims that children with parents who spend more time with them have greater human capital in terms of cognitive and non-cognitive outcomes. From that follows that children of more educated parents should have greater human capital as their parents spend more time with them. That may result in different education outcome.

When I control for those innate and acquired characteristics, I should obtain more precise results. The matching approach is a tool how to observe the pure effect of parental involvement since I control the baseline characteristics. I control the innate characteristics (e.g., IQ) by the covariates of the level of parental education. Assuming that the vector of covariates explains choice of parental involvement, matching approach will help identify the treatment effect. According to the results of matching method, the parental involvement do affect child's education outcome. Nevertheless, the results might be limited since I have restricted sample of Americans born in 1983 and 1984.

The surprising results of the limit setting variables might be caused not only by the limitations of the data. There is a chance that more problematic children have more limits. This "problematicity" is hidden in the error term but it might affect the probability of graduation from high school and at the same time it might affect limit setting variable. It may be a source of endogeneity. I suggest to include the variables of child's criminal behavior for further research.

## 5 Conclusion

The main goal of this study was to investigate the effect of parental involvement on children's education outcome, specifically, whether they completed high school or not. To observe such an effect, I used a sample from the National Longitudinal Survey of Youth 1997. The National

Longitudinal Survey of Youth 1997 is a publicly open database. The sample consisted of American respondents born in 1983 and 1984 who lived with both birth parents in a household when they were 12 to 14 years old. I investigated whether parental involvement in youth's life affects their graduation from high school. In addition, I researched on the impact of limit breaking by respondents on their education attainment.

The research results were conducted by the simple logistic regression and the matching method. The matching method - subclassification on propensity score - controls for the observable child's and parents' characteristics which can affect the outcome, i.e., the completion of high school, as well as the parental involvement variables.

In conclusion, there are observable significant effects in child's education outcome caused by parental involvement in child's life. The monitoring variables depict the increase in the expected probability of youth completing high school. Buchanan *et al.* (1992) claim that degree of monitoring has impact on child's scholastic achievement. The results support this statement. The research on the impact of limit breaking by respondents on their education attainment produces statistically significant effects. Moreover, I found surprising results for limit setting variables. The limit setting variables decrease the odds of completion of high school. This might be caused by limited sample or by exclusion of the variable describing criminal behavior of the respondents.

The results open door for possible further research. Firstly, it would be interesting to observe whether higher level of limit setting leads to higher level of limit breaking since "forbidden fruit is the sweetest". Also, for the observation of such an effect, it could be appropriate to include the variables for child's "problematicity". Secondly, according to Erford (1995) and Eccles *et al.* (1991), observing parallel questions is very useful. Discrepancies across the answers indicate misunderstanding in who in fact sets the limits, which often leads to limits breaking by youths from parents' point of view. It would be interesting to observe those discrepancies.

# References


Austin, P. C., P. Grootendorst, & G. M. Anderson (2007): "A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study." *Statistics in Medicine* **26(4)**: pp. 734–753.

Austin, P. C., M. M. Mamdani, T. A. Stukel, G. M. Anderson, & J. V. Tu (2005): "The use of the propensity score for estimating treatment effects: administrative versus clinical data." *Statistics in Medicine* **24(10)**: pp. 1563–1578.

Belfield, C. R. & H. M. Levin (2007): "The price we pay: economic and social consequences of inadequate education." *Washington, DC: The Brookings Institution* .

Bogenschneider, K. (1997): "Parental Involvement in Adolescent Schooling: A Proximal Process with Transcontextual Validity." *Journal of Marriage and Family* **59(3)**: pp. 718–733.

Bryan, J. (2005): "Fostering educational resilience and achievement in urban schools through school-family-community partnerships." *Professional School Counseling* **8**: p. 219.

Buchanan, C. M., E. E. Maccoby, & S. M. Dornbusch (1992): "Adolescents and their families after divorce: Three residential arrangements compared." *Journal of Research on Adolescence* **2(3)**: pp. 261–291.

Caliendo, M. & S. Kopeinig (2005): "Some Practical Guidance for the Implementation of Propensity Score Matching." *Discussion Paper No. 1588* .

Cole, S. A. (2017): *The Impact of Parental Involvement on Academic Achievement.* Ph.D. thesis, Northcentral University.

Eccles, J. S., C. M. Buchanan, C. Flanagan, A. Fuligni, C. Midgley, & D. Yee (1991): "Control Versus Autonomy During Early Adolescence." *Journal of Social Issues* **47(4)**: pp. 53–68.

Erford, B. T. (1995): "Parental Autonomy-Enhancement and Development of Self-Efficacy." *Psychological Reports* **77(3)**: pp. 1347–1353.

Feinstein, L. & J. Symons (1999): "Attainment in Secondary School." *Oxford Economic Papers* .

Guryan, J., E. Hurst, & M. Kearney (2008): "Parental Education and Parental Time with Children." *Journal of Economic Perspectives* **22(3)**: pp. 23–46.

Hoover-Dempsey, K. V. & H. M. Sandler (1997): "Why Do Parents Become Involved in TheirChildren's Education? ." *Vanderbilt University* **67(1)**: pp. 3–42.

Kalil, A., R. Ryan, & M. Corey (2012): "Diverging Destinies: Maternal Education and the Developmental Gradient in Time with Children." *Demography* **49(4)**: pp. 1361–1383.

McCallumore, K. M. & E. F. Sparapani (2010): "The importance of the ninth grade on high school graduation rates and student success in high school." *Education* **130(3)**.

Mcdowell, K., A. Jack, & M. Compton (2018): "Parent Involvement in Pre-Kindergarten and the Effects on Student Achievement." *The Advocate* **23**.

McLanahan, S. (2004): "Diverging destinies: How children are faring under the second demographic transition." *Demography* **41**: pp. 607–627.

Morin, A. (2020): "How to discipline kids with positive and negative consequences." *Verywell Family* .

Morin, A. (2021): "Reasons why it is important to set limits with kids." *Verywell Family* .

Neidell, M. J. (2000): "Early Parental Time Investments In Children's Human Capital Development: Effects Of Time In The First Year On Cognitive And Non-Cognitive Outcomes." *UCLA Economics Working Papers 806, UCLA Department of Economics* .

Orihara, S. & E. Hamada (2021): "Determination of the optimal number of strata for propensity score subclassification." *Statistics & Probability Letters* **168**: p. 108951.

Rasmussen, A. W. (2009): "Allocation of Parental Time and the Long-Term Effect on Children's Education." *Department of Economics and CIM, Aarhus University.* **Working paper 09-22**.

Rosenbaum, P. R. & D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects." *Biometrika* **70(1)**: pp. 41–55.

Sum, A., I. Khatiwada, & J. McLaughlin (2009): "The consequences of dropping out of high school: joblessness and jailing for high school dropouts and the high cost for taxpayers." *Boston: Center for Labor Market Studies Publications* **(23)**.

Zaff, J. F., A. Donlan, A. Gunning, S. E. Anderson, E. McDermott, & M. Sedaca (2017): "Factors that Promote High School Graduation: a Review of the Literature." *Educ Psychol Rev* **29**: pp. 447–476.

# Appendix

Figure A1: Love plot - Monitoring by mother

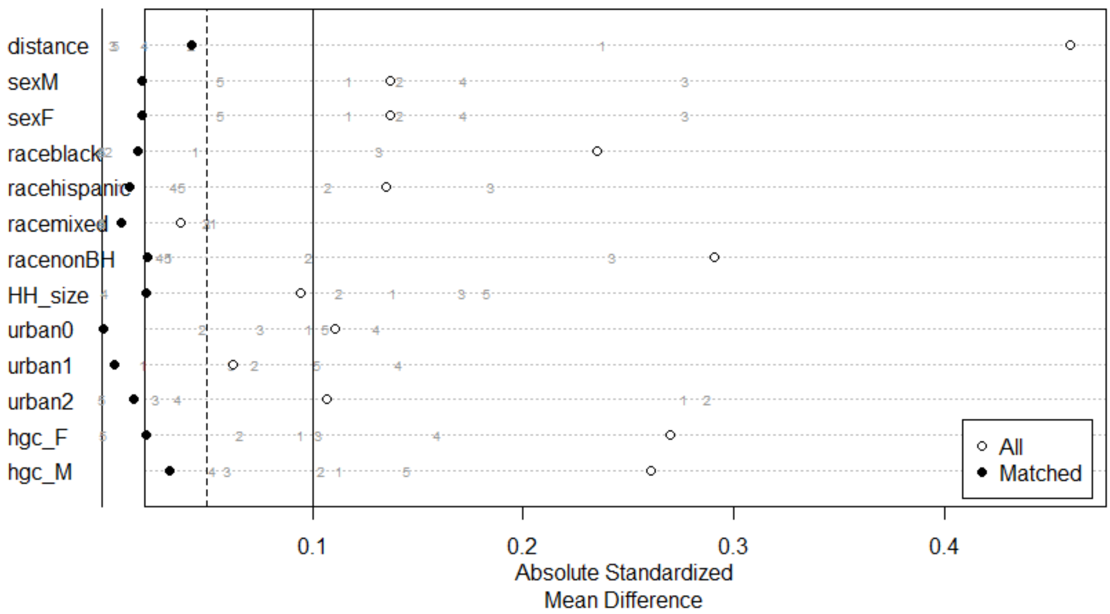



Figure A2: Monitoring by mother

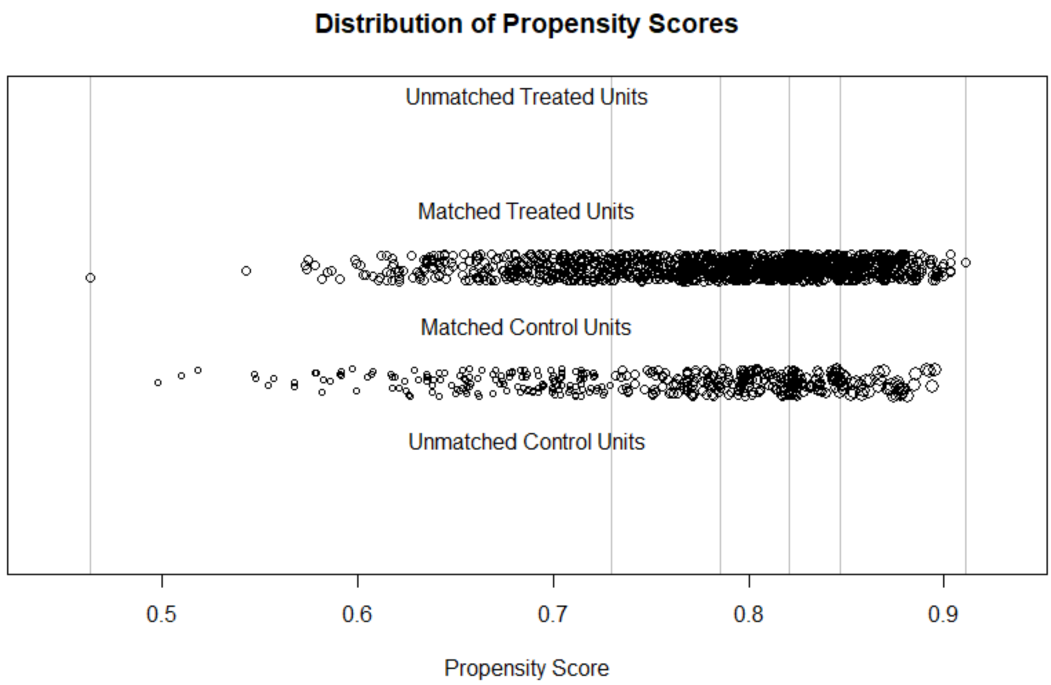


Figure A3: Propensity Score Histograms - Monitoring by mother

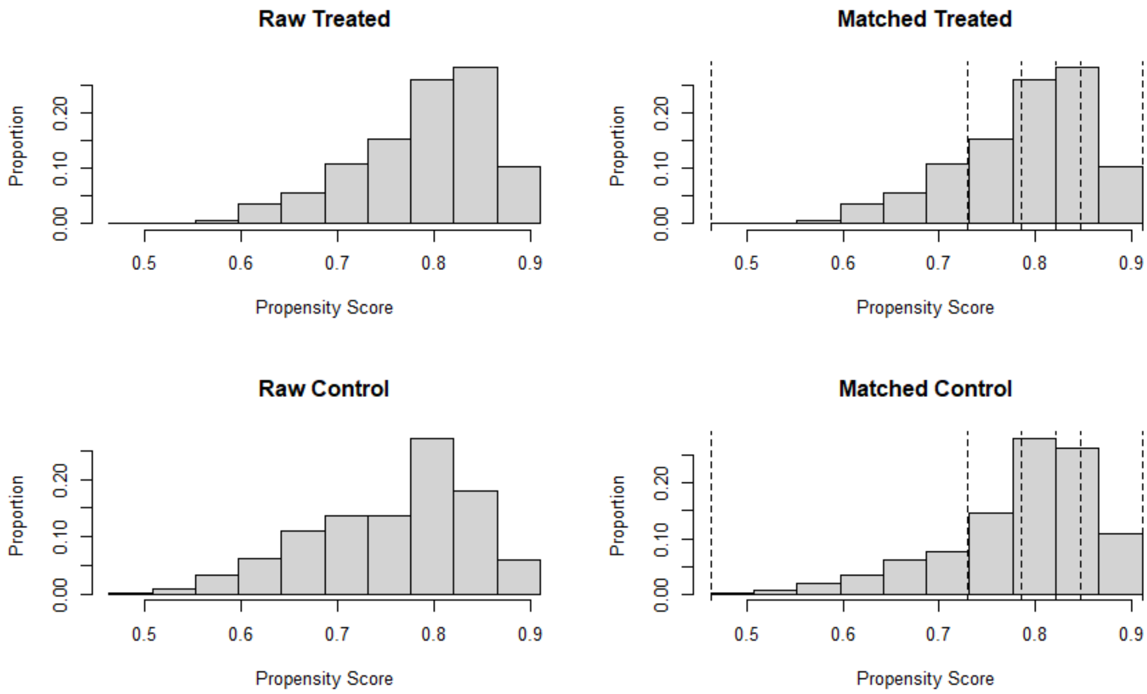



Figure A4: Love plot - Monitoring by father

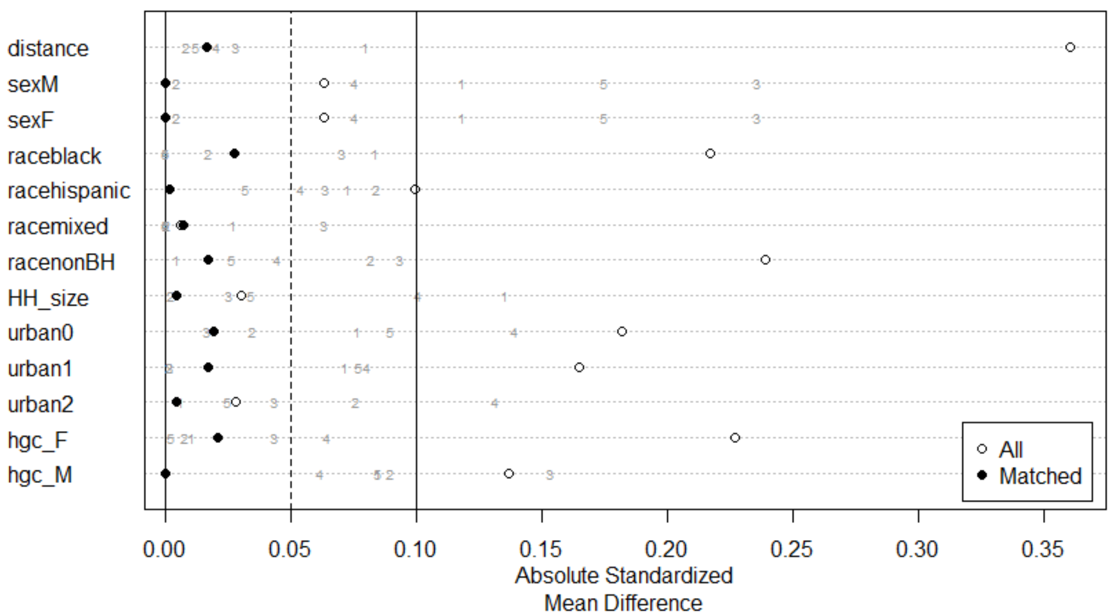


Figure A5: Monitoring by father

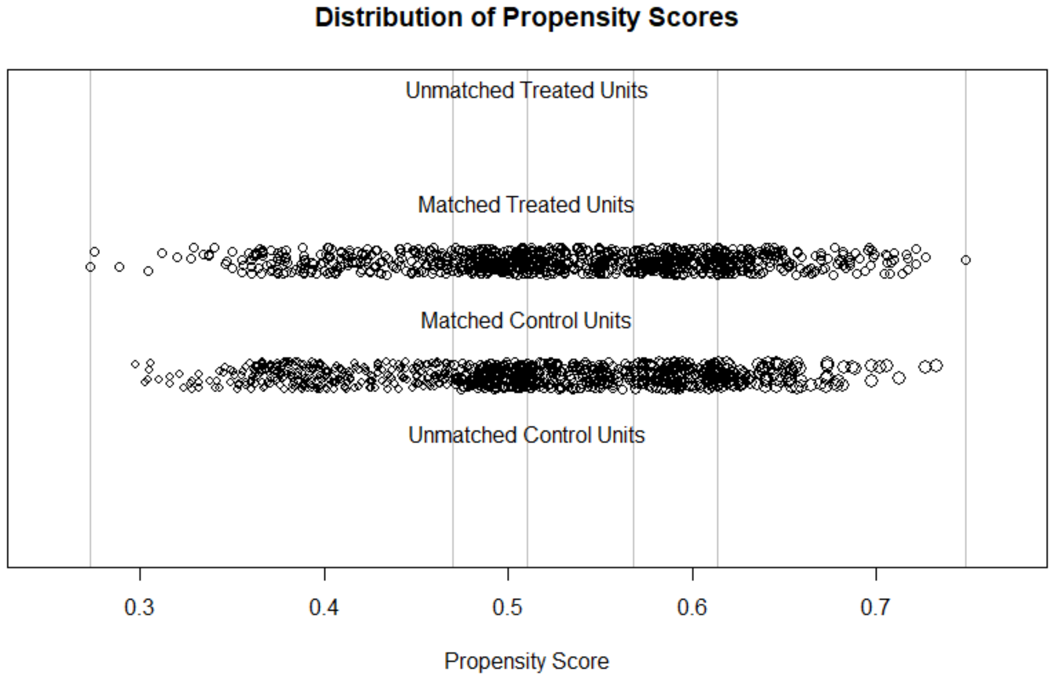



Figure A6: Propensity Score Histograms - Monitoring by father

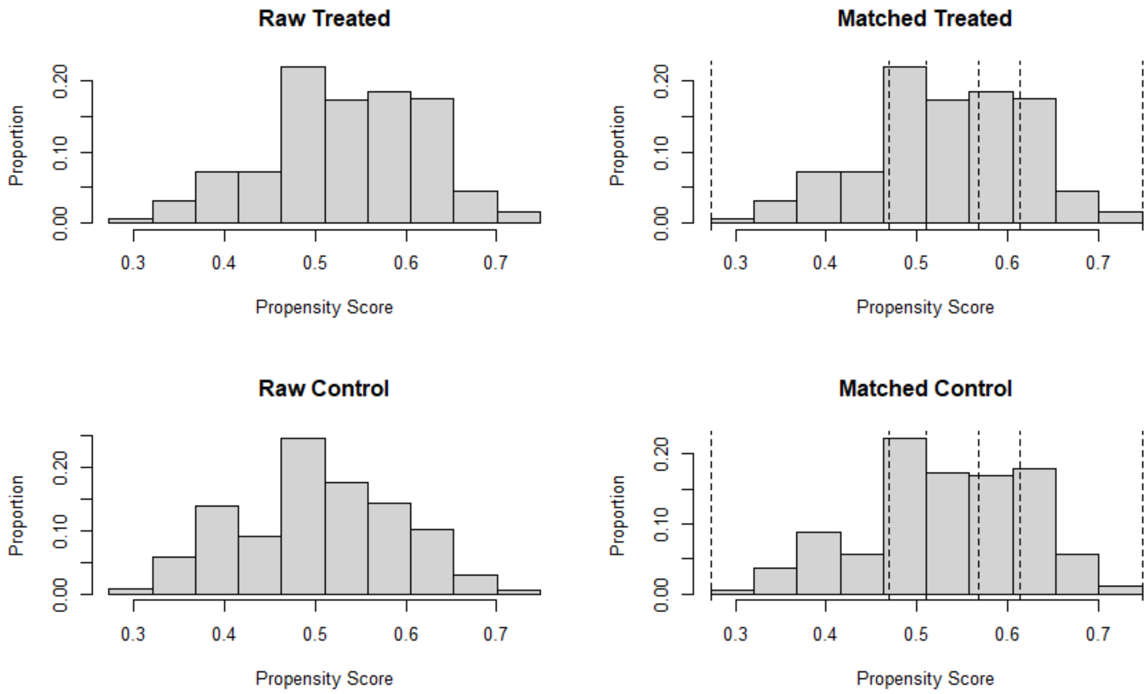


Figure A7: Love plot - Limit setting (youth)

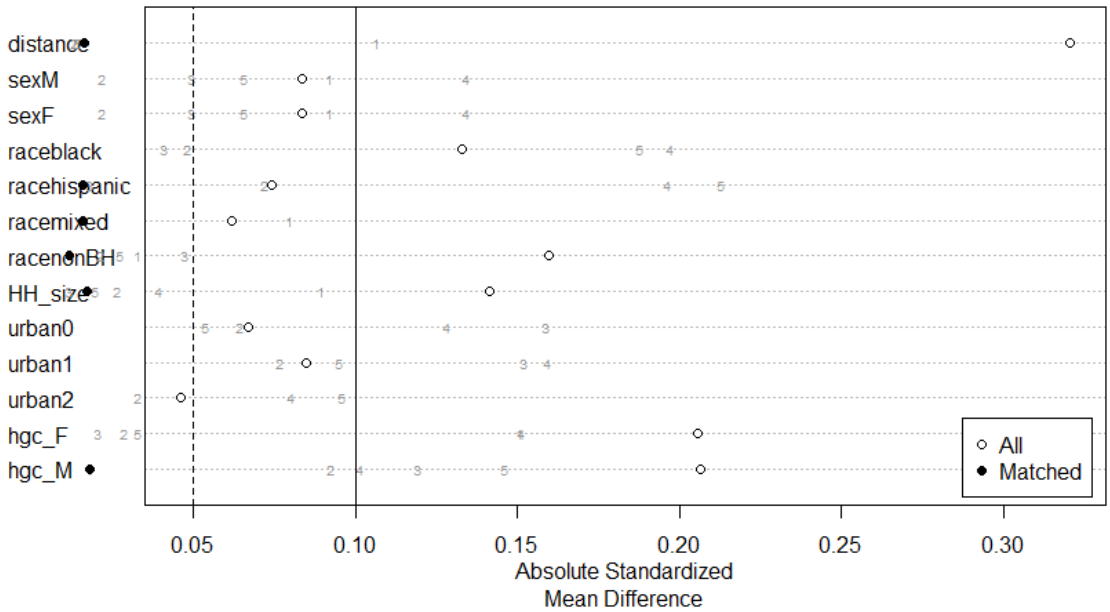



Figure A8: Limit setting (youth)

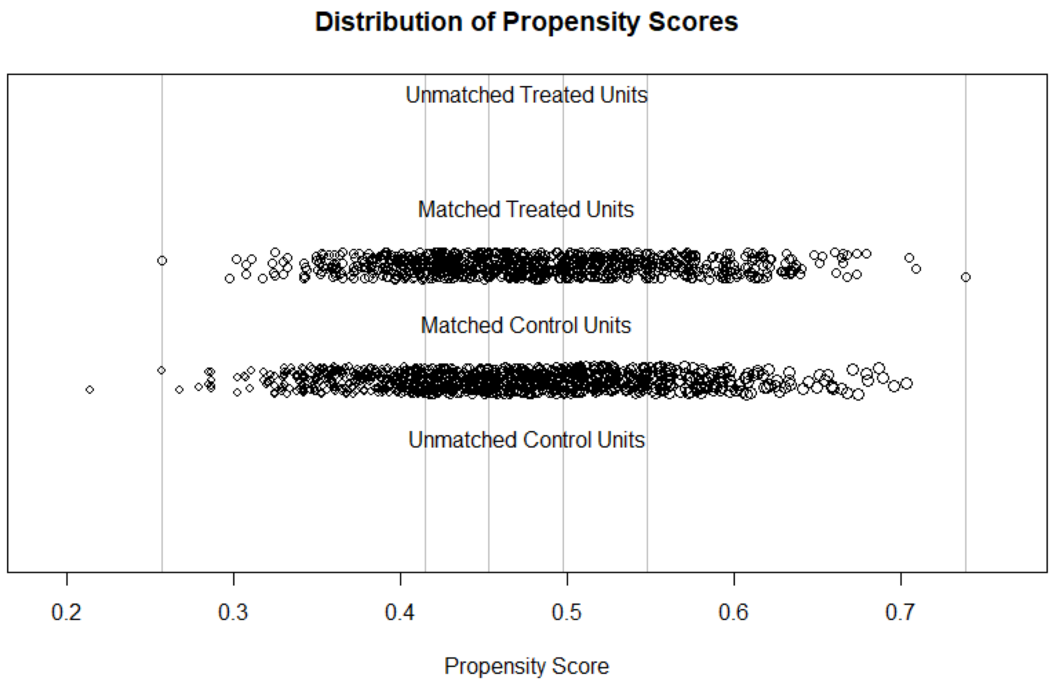


Figure A9: Propensity Score Histograms - Limit setting (youth)

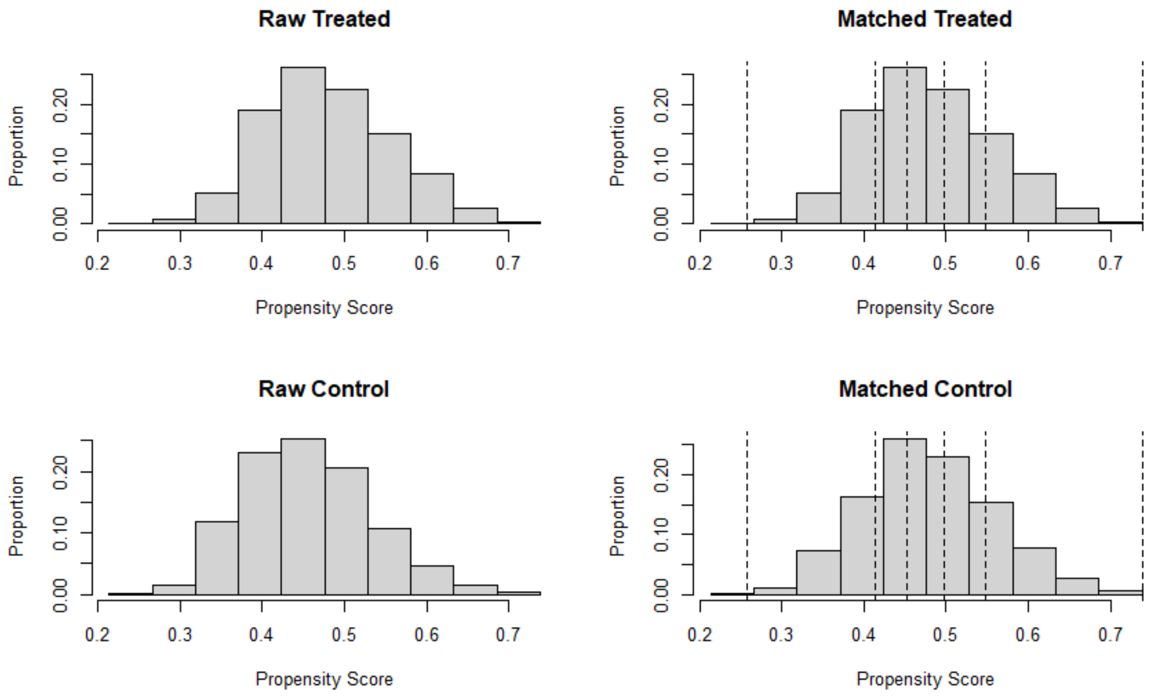



Figure A10: Love plot - Limit setting (parent)

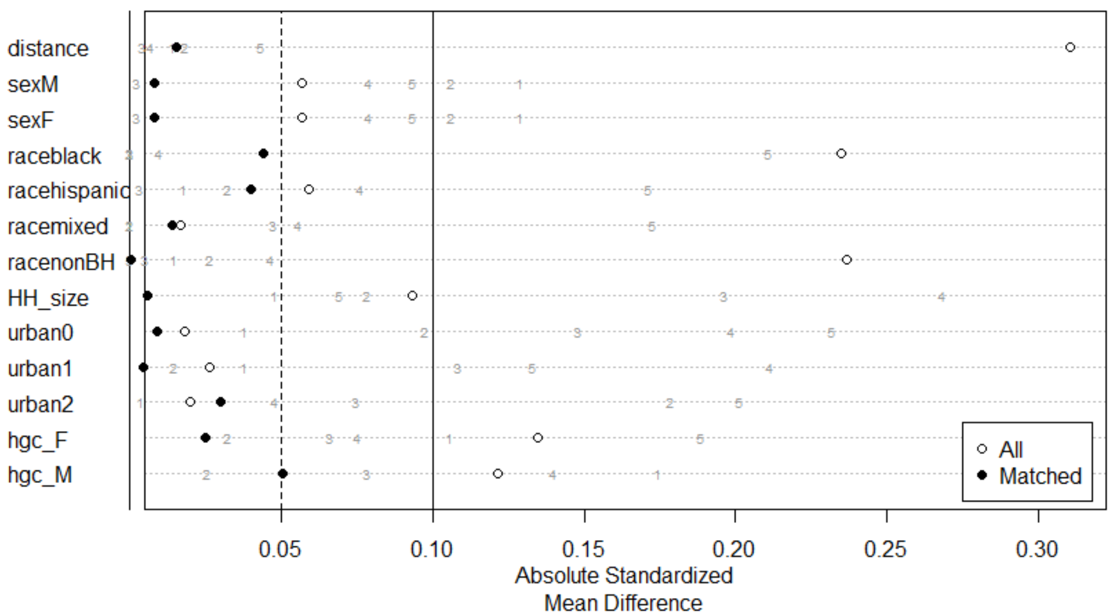


Figure A11: Limit setting (parent)

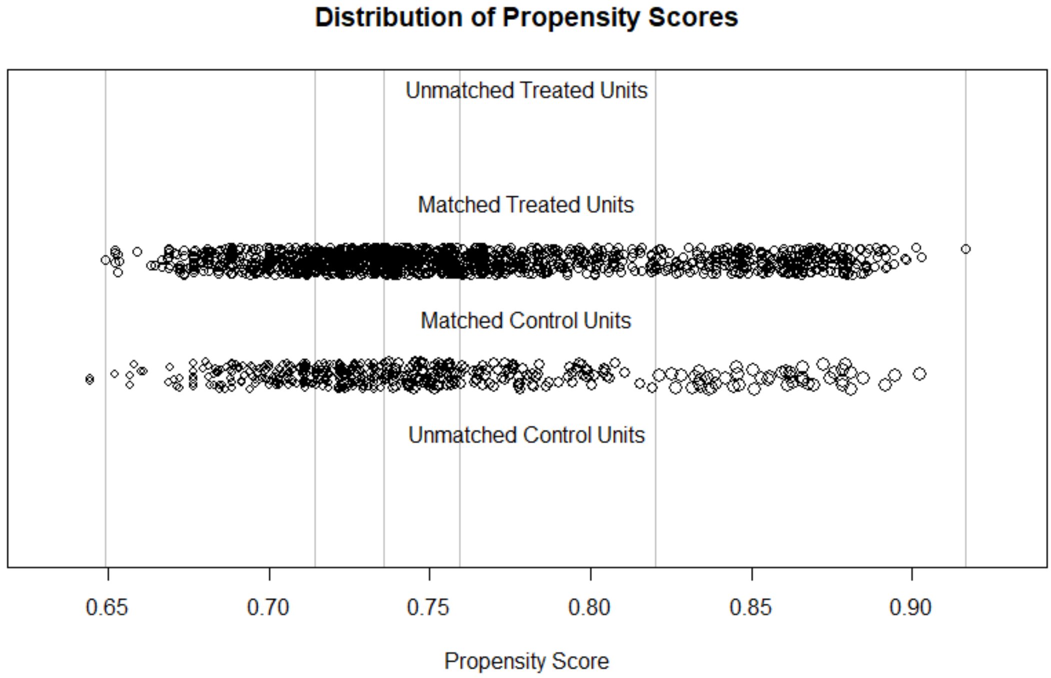



Figure A12: Propensity Score Histograms - Limit setting (parent)

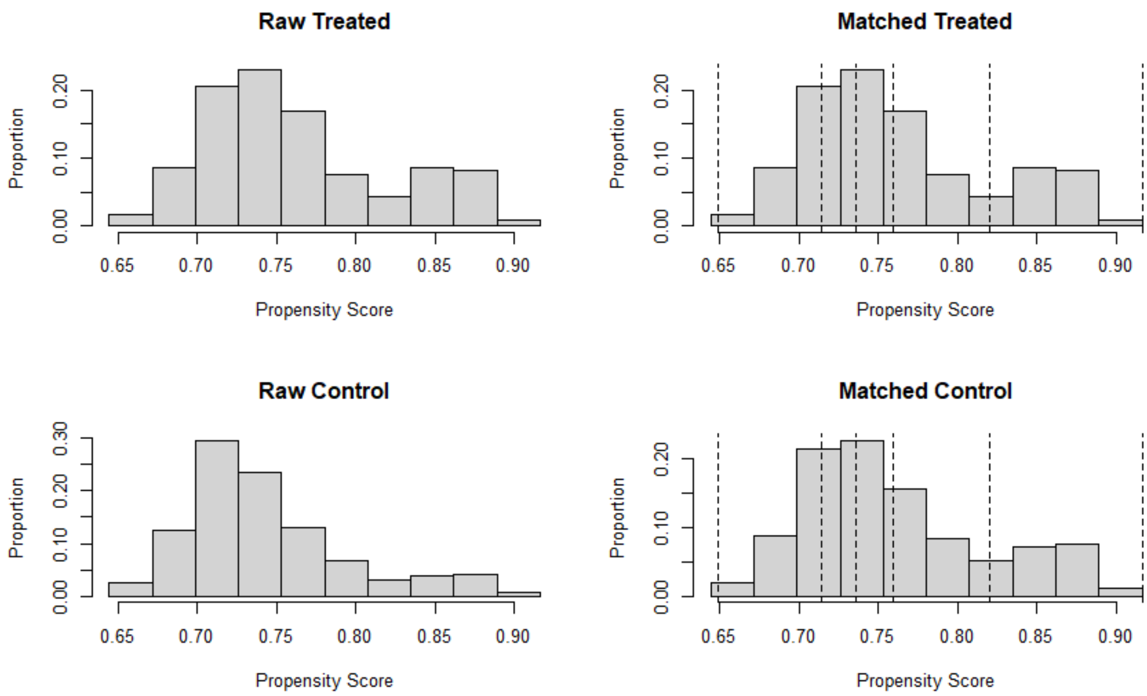


Figure A13: Love plot - Limit breaking (youth)

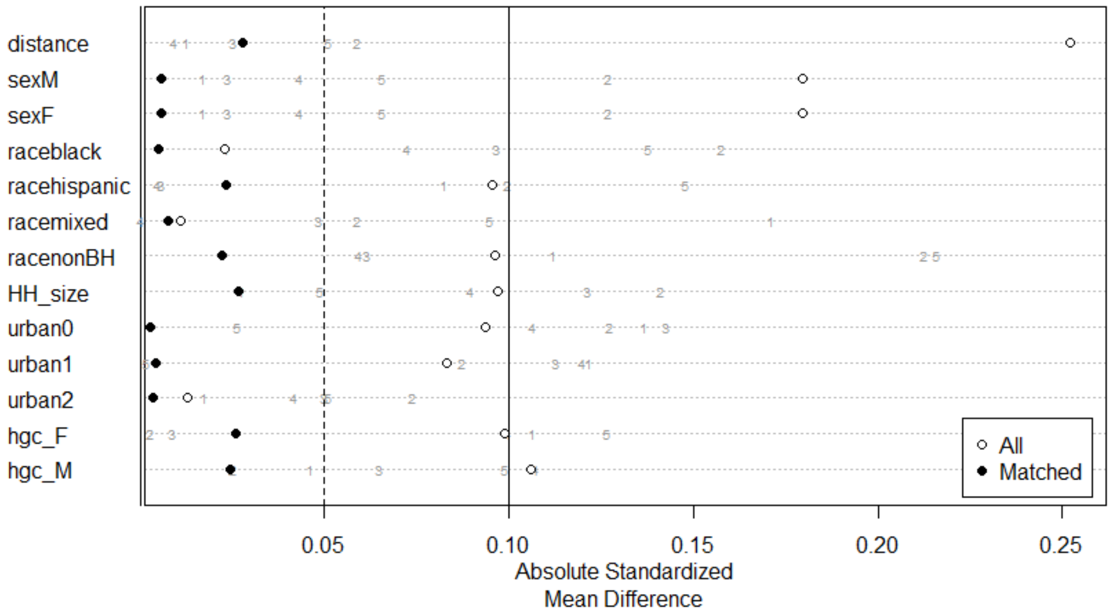



Figure A14: Limit breaking (youth)

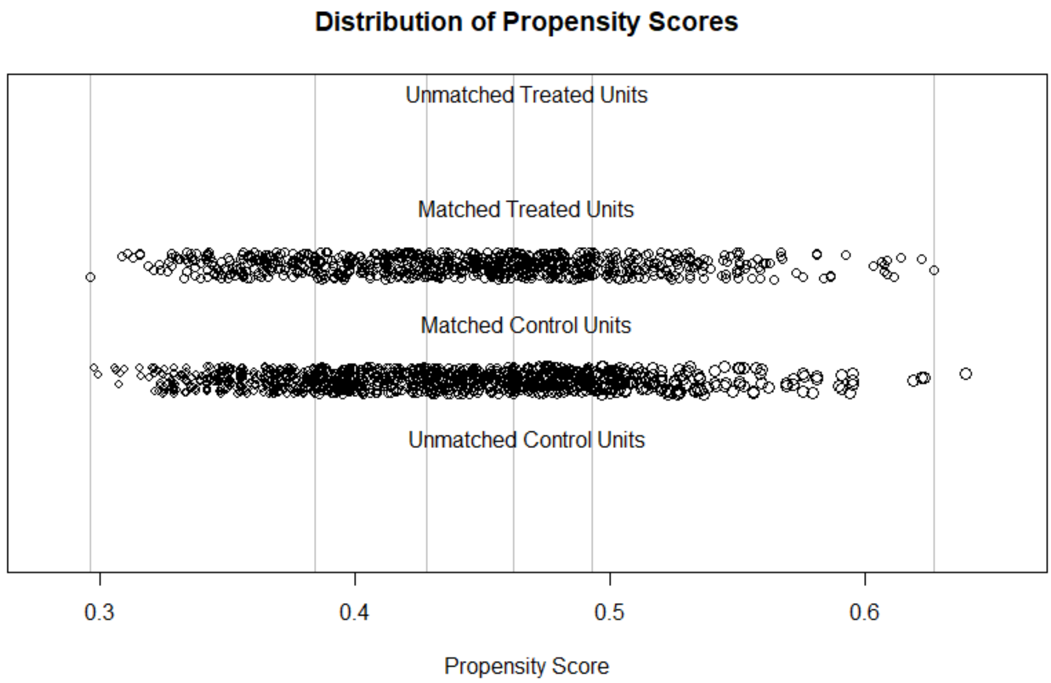


Figure A15: Propensity Score Histograms - Limit breaking (youth)

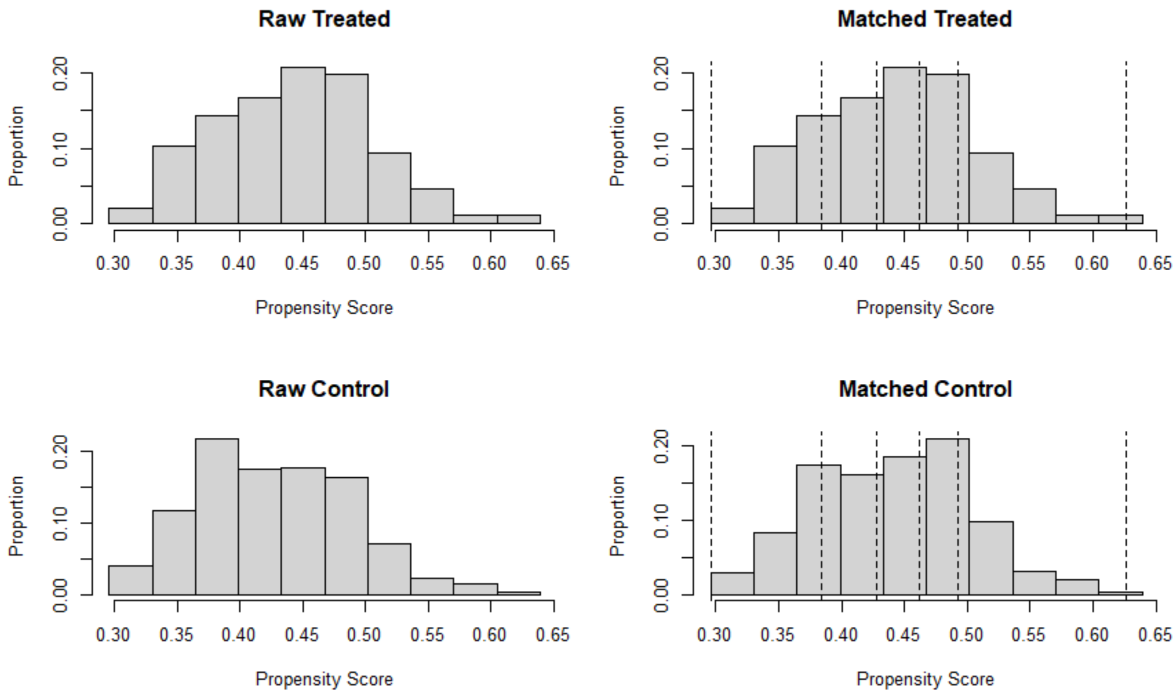



Figure A16: Love plot - Limit breaking (parent)

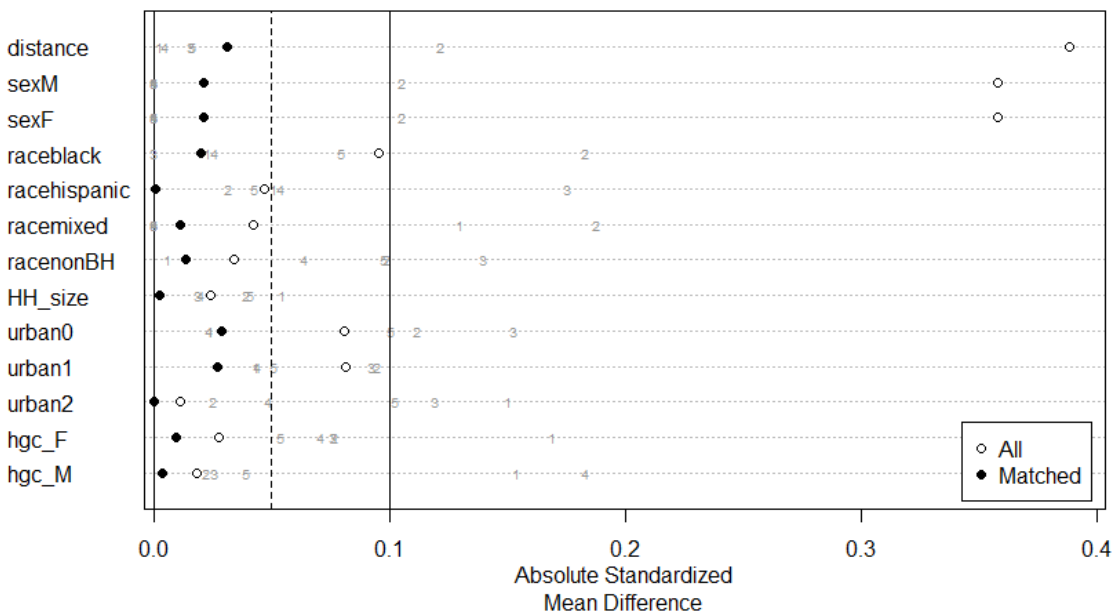


Figure A17: Limit breaking (parent)

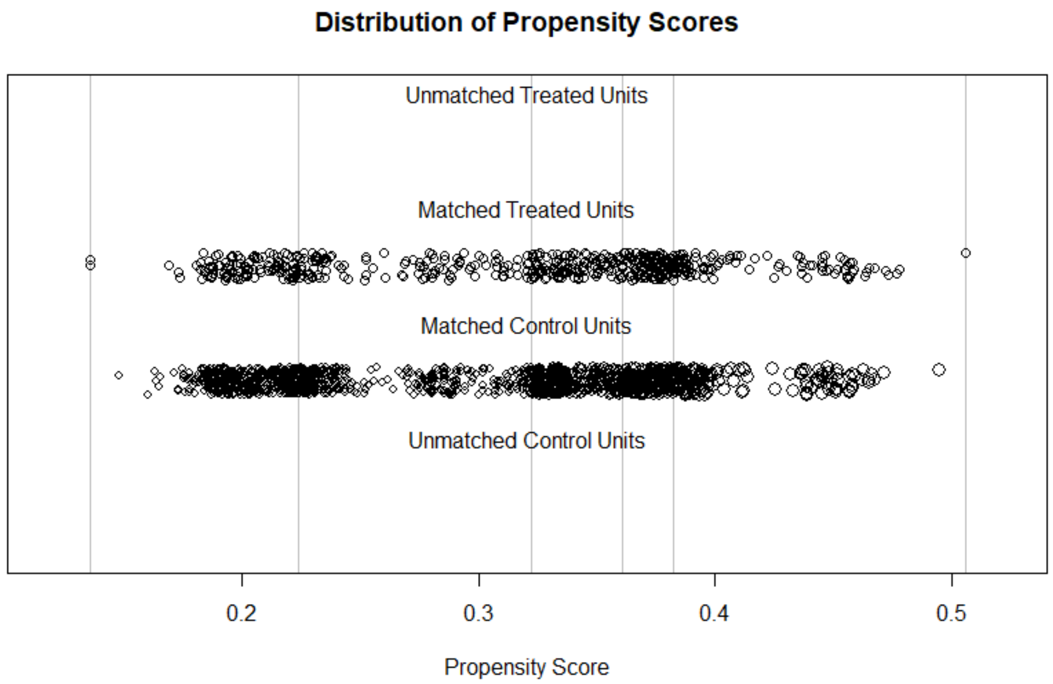
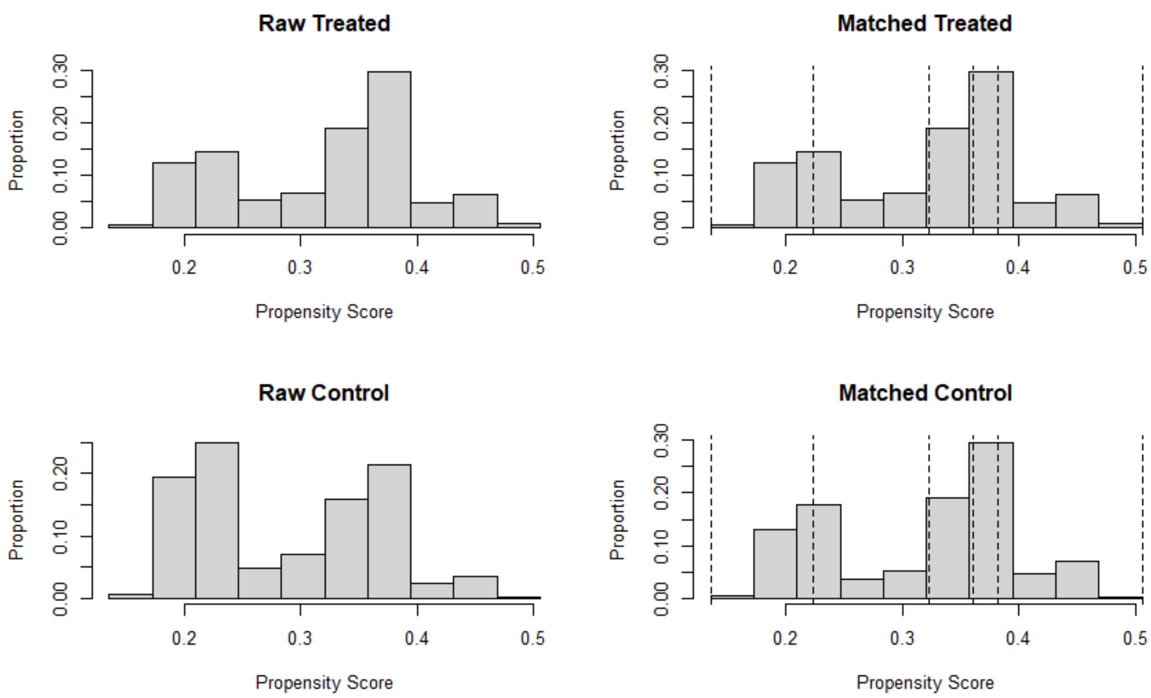



Figure A18: Propensity Score Histograms - Limit breaking (parent)



**Correlation between education attainment variables** Among the respondents, 84.4 % have graduated from high school (as is known from Table 1). Concerning the sex, 87.4 % of females and 81.6 % of males graduated from high school. Those portions are comparable to those of their parents (88.0 % for mother and 80.1 % for fathers). To understand the associations between variables, I design the contingency table. Moreover, I derive the odds ratio and phi-coefficient to interpret the strength of association between examined variables. The following contingency table displays the multivariate frequency distribution of the variables. I use binary variables whether the respondents and their parents graduated from high school or not.

Table A1: Contingency table

| | **HS M** | | | **HS F** | | |
|---|---|---|---|---|---|---|
| | no | yes | total | yes | no | total |
| **HS respondent** | | | | | | |
| no | 113 | 129 | 292 | 121 | 171 | 292 |
| yes | 224 | 1,352 | 1,579 | 251 | 1,325 | 1,579 |
| total | 337 | 1,531 | | 372 | 1,496 | |
| | Odds ratio: 3.81 | | | Odds ratio: 3.73 | | |
| | Phi coef.: 0.231 | | | Phi coef.: 0.232 | | |

*Notes*: HS respondent - graduation from high school by respondents, HS M - graduation from high school by mothers, HS F - graduation from high school by fathers.

The cells of the table give the counts of respondents that share each combination of high school graduation. Based on the odds ratios and phi-coefficients, there is a positive association in the variables, i.e., in respondents' and mothers' as well as fathers' graduation from high school.

Table A2: Mean & SD of parental involvement variables

| | Male | | Female | |
|---|---|---|---|---|
| | 1 | 0 | 1 | 0 |
| **Mon M** | | | | |
| HH size | 4.66 | 4.75 | 4.73 | 4.89 |
| | (1.18) | (1.36) | (1.34) | (1.52) |
| hgc M | 13.19 | 12.48 | 12.92 | 12.05 |
| | (2.94) | (2.98) | (2.93) | (3.15) |
| hgc F | 13.13 | 12.07 | 12.95 | 12.31 |
| | (3.24) | (3.30) | (3.21) | (3.41) |
| **Mon F** | | | | |
| HH size | 4.70 | 4.67 | 4.74 | 4.78 |
| | (1.15) | (1.32) | (1.46) | (1.30) |
| hgc M | 13.21 | 12.80 | 12.38 | 13.05 |
| | (2.95) | (2.97) | (2.93) | (3.04) |
| hgc F | 13.24 | 12.46 | 13.18 | 12.46 |
| | (3.41) | (3.08) | (3.21) | (3.27) |
| **Lim Y** | | | | |
| HH size | 4.74 | 4.62 | 4.92 | 4.63 |
| | (1.29) | (1.17) | (1.65) | (1.53) |
| hgc M | 12.72 | 13.31 | 12.47 | 13.03 |
| | (2.96) | (2.94) | (2.99) | (2.95) |
| hgc F | 12.59 | 13.14 | 12.40 | 13.17 |
| | (3.24) | (3.30) | (3.06) | (3.37) |
| **Lim P** | | | | |
| HH size | 4.71 | 4.57 | 4.79 | 4.68 |
| | (1.28) | (1.06) | (1.38) | (1.38) |
| hgc M | 12.93 | 13.31 | 12.66 | 13.02 |
| | (2.96) | (2.96) | (2.91) | (3.21) |
| hgc F | 12.76 | 13.27 | 12.73 | 13.10 |
| | (3.25) | (3.38) | (3.26) | (3.26) |
| **Broke Y** | | | | |
| HH size | 4.71 | 4.65 | 4.90 | 4.67 |
| | (1.23) | (1.23) | (1.47) | (1.31) |
| hgc M | 12.90 | 13.13 | 12.45 | 12.95 |
| | (3.09) | (2.85) | (3.17) | (2.85) |
| hgc F | 12.71 | 13.03 | 12.59 | 12.97 |
| | (3.33) | (3.24) | (3.55) | (3.05) |
| **Broke P** | | | | |
| HH size | 4.72 | 4.66 | 4.78 | 4.75 |
| | (1.31) | (1.18) | (1.49) | (1.35) |
| hgc M | 12.90 | 13.09 | 12.98 | 12.69 |
| | (2.89) | (3.01) | (2.58) | (3.09) |
| hgc F | 13.03 | 12.60 | 13.13 | 12.74 |
| | (3.06) | (3.40) | (3.14) | (3.29) |

*Notes:* Mon M - monitoring by mother; Mon F - monitoring by father; Lim Y - limit setting reported by youth; Lim P - limit setting reported by parent; Broke Y - limit breaking reported by youth; Broke P - limit breaking reported by parent. HH size - household size (range: $<2;14>$); hgc M - highest grade completed by mother (range: $<2;20>$); hgc F - highest grade completed by father (range: $<2;20>$).

Table A3: Control variables (logistic regression)

| | $Mon_M$ | $Mon_F$ | $Lim_Y$ | $Lim_Y$ | $Broke_Y$ | $Broke_Y$ |
|---|---|---|---|---|---|---|
| **Control variables**: | | | | | | |
| sexF | $0.06^{***}$ | $0.06^{***}$ | $0.06^{***}$ | $0.06^{***}$ | $0.06^{***}$ | $0.05^{**}$ |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| racehispanic | $0.05^{**}$ | $0.06^{**}$ | $0.06^{**}$ | $0.06^{**}$ | $0.06^{**}$ | $0.05^{**}$ |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| racemixed | $-0.05$ | $-0.06$ | $-0.06$ | $-0.06$ | $-0.06$ | $-0.07$ |
| | (0.13) | (0.13) | (0.13) | (0.13) | (0.13) | (0.13) |
| racenonBH | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| HHsize | $-0.00$ | $-0.00$ | 0.00 | 0.00 | 0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| urban1 | 0.00 | 0.00 | $-0.00$ | $-0.00$ | 0.00 | 0.00 |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| urban2 | $-0.02$ | $-0.02$ | $-0.03$ | $-0.03$ | $-0.02$ | $-0.02$ |
| | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) |
| hgcF | $0.02^{***}$ | $0.02^{***}$ | $0.02^{***}$ | $0.02^{***}$ | $0.02^{***}$ | $0.02^{***}$ |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| hgcM | $0.01^{***}$ | $0.01^{***}$ | $0.01^{***}$ | $0.01^{***}$ | $0.01^{***}$ | $0.01^{***}$ |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Num. obs. | 1868 | 1868 | 1868 | 1868 | 1868 | 1868 |
| Log Likelihood | $-738.30$ | $-740.55$ | $-743.02$ | $-740.74$ | $-736.46$ | $-735.77$ |
| Deviance | 1476.59 | 1481.10 | 1486.03 | 1481.48 | 1472.92 | 1471.54 |
| AIC | 1498.59 | 1503.10 | 1508.03 | 1503.48 | 1494.92 | 1493.54 |
| BIC | 1559.45 | 1563.96 | 1568.89 | 1564.33 | 1555.78 | 1554.40 |

*Notes:* $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Coefficient (marginal effect dy/dx) for categorical variable is the discrete change from the base. Robust standard errors are provided in parentheses, sexF - female; racehispanic - Hispanic respondents; racemixed - mixed race respondents; racenonBH - Non-Black/Non-Hispanic; HH size - household size; urban1 - urban; urban2 - unknown; hgcF - highest grade completed by father; hgcM - highest grade completed by mother.

Table A4: Model statistics (subclassification)

| | $Mon_M$ | $Mon_F$ | $Lim_Y$ | $Lim_Y$ | $Broke_Y$ | $Broke_Y$ |
|---|---|---|---|---|---|---|
| Num. obs. | 1868 | 1868 | 1868 | 1868 | 1868 | 1868 |
| Log Likelihood | $-798.40$ | $-804.39$ | $-807.62$ | $-805.66$ | $-799.15$ | $-800.36$ |
| Deviance | 1596.80 | 1608.78 | 1615.24 | 1611.32 | 1598.30 | 1600.72 |
| AIC | 1600.80 | 1612.78 | 1619.24 | 1615.32 | 1602.30 | 1604.72 |
| BIC | 1611.86 | 1623.85 | 1630.30 | 1626.39 | 1613.36 | 1615.78 |

*Notes:* $Mon_M$ - monitoring by mother; $Mon_F$ - monitoring by father; $Lim_Y$ - limit setting reported by youth; $Lim_P$ - limit setting reported by parent; $Broke_Y$ - limit breaking reported by youth; $Broke_P$ - limit breaking reported by parent.

# IES Working Paper Series

## 2022

1. Klara Kantova: *Parental Involvement and Education Outcomes of Their Children*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz