What explains the gender gap in high school dropout rates? Experimental and administrative evidence

Ingvild Almås   Alexander W. Cappelen   Kjell G. Salvanes
Erik Ø. Sørensen   Bertil Tungodden*

December 29, 2015

Abstract

A new gender gap has emerged in many societies: males lag behind in terms of educational outcomes. In a largely representative sample of Norwegian adolescents, we examine why males make different school choices than females and why they are more likely to drop out of high school. We exploit a unique data set, combining rich experimental data with high-quality administrative data, and study how family background and personal characteristics explain dropout rates. The paper provides three main findings. First, we show that both family background and personal characteristics are of great importance in explaining dropout, but they do not account for the observed gender difference. Second, we show that the gender difference in dropout rates appears both when the adolescents select into the college track and after they have started: girls are much more likely to choose the college track and to continue after they have started. Third, we show that very different processes guide the choices of the boys and the girls of whether to drop out from the college track. We argue that the findings have important

*Department of Economics, NHH Norwegian School of Economics, 5045 Bergen, Norway. Emails: ingvild.almas@nhh.no, alexander.cappelen@nhh.no, kjell.salvanes@nhh.no, erik.sorensen@nhh.no, bertil.tungodden@nhh.no. We would like to thank the municipal school authority in Bergen for their cooperation; Kristin Risvand Mo for administrative support; Atle Askeland, Bjørn Ivar Grøttå, and Sarah Marie Søiland for IT-support; and Eli Birgitte S. Bergmark, Espen Bolhaug, Andreas Tufteland Engelsen, Kasper Thorin Fellkjaer, Sebastian Fest, Ingar Haaland, Sigve Garsjø, Vi-Vi Ton Nu, Kristin Orset, Henrik Reigstad, Mari Sakkestad, Johanne Amundsen Vik, Kristina Øystese, and Silje Åsnes for research assistance. The project was financed by support from the Research Council of Norway, research grant 185831, and administered by The Choice Lab, Norwegian School of Economics.
implications for how society addresses the problem of adolescents dropping out from the college track.

1 Introduction

In most developed countries, girls outperform boys in primary school through college (OECD, 2012). Notably, the reversal of the gender gap took place at the same period across countries; around 1970 the high school completion rate and college attendance was higher for boys, while today it is in favor of girls. What is behind this is not well understood, but a recent wave of papers suggest that both family background and non-cognitive skills are important for explaining gender differences in school choices and school performance (Heckman et al., 2013; Bertrand and Pan, 2013; Buser et al., 2014; Autor and Wasserman, 2015).

We exploit a unique data set, combining rich experimental data with high-quality administrative data, to study dropout from the college track at high school in Norway, and, in particular, why boys are more likely to drop out than girls. A representative sample of Norwegian adolescents took part in a lab experiment in 9th grade (14-15 years old), just before they were to apply to high school and, in collaboration with Statistics Norway, we matched the data from the experiment to Norwegian register data on family background and school choices. The register data provide us with information about whether the adolescents chose the college track at high school, and if they did, whether they have stayed on the college track two years after they started.

We focus on overall dropout from the college track (which includes both adolescents who did not choose the college track and adolescents who dropped out after starting the college track), since it is well demonstrated that college attainment is an increasingly important determinant of success in the labor market and also more broadly strongly associated with a wide range of positive life outcomes (Autor et al., 2008; Oreopoulos and Salvanes, 2011). In Figure 1, we show that the gender difference in college attainment in Norway is very much in line with what has been observed internationally (Murnane, 2013): boys are increasingly lagging behind. In recent years, boys in Norway are almost 30% less likely than girls to have a college education at the age of 35. The same pattern emerges in our

---

1The college track is the academically oriented track at high school and is required for university studies. The vocational track does not exclude the possibility of pursuing some further studies after high school, but severely limits the set of educational opportunities.

2We define a person to have had some college education if he or she has reached NUS-level 5 in the Norwegian classification system of education. Most of the studies at level 5 require that the person has followed the college track at high school. For more details on the Norwegian educational system, see Liu et al. (2014).
sample, as shown in Figure 2: the overall dropout rate from the college track for the boys is 30% higher than for the girls: 45.1% of girls and 58.7% of boys in our sample have dropped out from the college track at high school, which amounts to a gender difference of almost 14 percentage points.

Our analysis contains two main parts. First, we study the extent to which family background and personal characteristics (including ability, preferences, beliefs, and non-cognitive skills) can explain overall dropout from the college track, and, in particular, the observed gender difference. The second part of the analysis studies in more detail the two choices driving overall dropout: the choice of college track at high school and the choice of staying in the college track after having started.

The paper provides three main findings. First, we show that both family background and personal characteristics are of great importance in explaining dropout, but they do not account for the observed gender difference. In fact, we find that the estimated male effect on dropout rates increases when we control for abilities, preferences, and beliefs. Second, we show that the gender difference in dropout rates appears both when the adolescents select into the college track and after they have started: girls are much more likely to choose the college track and stay on the track after they have started. Third, we show that very different processes guide the choices of boys and girls of whether to drop out from the college track.

The paper contributes to the literature focusing on the growing public concern for the high dropout rates, and the "boy crisis" in school (Goldin et al., 2006; Murrian, 2013; Bertrand and Pan, 2013; Figlio et al., Forthcoming; Fortin et al., Forthcoming; Baker and Milligan, 2013). In a recent wave of papers, non-cognitive skills such as emotional and behavioral factors are found to explain school performance, and differences across these dimensions are also related to gender differences in school performance (Heckman et al., 2013), in particular, boys school performance is affected by their disruptive behavior (Bertrand and Pan, 2013). We contribute to this literature by shedding light on the importance of family background and a rich set of personal characteristics that are potentially powerful in explaining dropout rates, and also by providing suggestive evidence of how family background may contribute to shape children’s and adolescents’ behavior (Alm˚as et al., Forthcoming; Cunha and Heckman, 2007; Autor et al., 2015). We further relate to the recent literature in experimental economics that studies the association between the willingness to compete and school choices (Buser et al., 2014; Zhang, 2012), and the more general literature on gender differences in preferences (Croson and Gneezy, 2009; Niederle and Vesterlund, 2011).

3It should be noted that the college track is open and free of charge for all students in Norway, independent of their grades from secondary school.
2 Sample and experimental design

Our sample consists of 9th grade students (14-15 years old) in Norway when the lab experiment was conducted in 2011. They were recruited from randomly selected schools in Bergen municipality, which is largely representative for Norway. At each school, we randomly selected some classes, and all students in these classes received a personal invitation to participate in the experiment. Out of 603 invited students from nine schools, 523 took part in the experiment, giving us a mean participation rate of 87%. Since the 9th grade is compulsory in Norway, with no grade repetition and basically all students complete the grade, we consider the sample to be representative for this age group in Norway. In collaboration with Statistics Norway, we matched the data from the experiment to Norwegian register data, which is a linked national administrative high quality data set. We have detailed parental background information on education and income and data on school choices for 483 of the 523 adolescents. In the present analysis, the administrative data provide us with indicator variables for whether at least one of the parents has completed some college and whether the adolescent chose the college track at high school and has continued on the college track two years after they started.

We conducted ten experimental sessions at NHH Norwegian School of Economics, where each session lasted for approximately two hours and used a web-based interface. All students received a show-up fee of 50 NOK (approximately 8 USD), in addition to what they earned in the lab experiment. The participants were not given any feedback on the different incentivized parts of the experiment until the end of the session. They were then given an overview of the outcomes and paid the sum of what they had earned in each part. The average total payment from the experiment was 361 NOK. The experiment was double blind, i.e., neither participants nor experimenters could associate decisions with particular participants.

The experimental session consisted of two parts, an incentivized part and a non-incentivized part. In the present analysis, we focus on the following components measures in the incentivized part: performance on a math task as a proxy for ability, competition preferences, risk preferences, time preferences, and the participants’ beliefs about own abilities and about earnings in the labor market. The experiment also included a nonincentivized part, where, among other things, we collected data on psychological traits using the Big Five Inventory (John et al.).

---

4This is confirmed by comparing family background data for our sample with national data. The distribution of income and education of the parents of the participants in our sample is in line with official statistics for Norway. For a more detailed discussion of this issue, see Almås et al. (Forthcoming).

5A detailed description of the experiment is reported in Almås et al. (Forthcoming).
3 Analysis

Our analysis proceeds in three steps. First, we provide descriptive statistics on the extent to which boys and girls differ in terms of family background and personal characteristics. Second, we study how overall dropout rates relate to these explanatory variables, and third we analyze the underlying mechanisms.

3.1 Descriptive statistics

Figure 4 provides an overview of how the girls and boys differ in the explanatory variables. As expected, we do not observe any gender difference with respect to the likelihood of the adolescent coming from a family where none of the parents has some college education (boys: 38.2%, girls: 36.8%). Family background can thus only explain the gender difference in dropout rates if it affects boys and girls differently. In terms of personal characteristics, however, we do indeed observe some important gender differences. Boys perform better on the ability test, they are more competitive (but in our sample not more risk-taking), they are more patient, and more confident regarding their own ability. Girls, on the other hand, have more informed beliefs about the earnings in the labor market. The girls and boys also differ importantly when it comes to non-cognitive skills as measured by the Big Five. The girls are more agreeable, conscientious, extrovert and open, but also strikingly more neurotic. Overall, the observed gender differences in personal characteristics might indeed contribute to explain the observed gender difference in dropout rates. For example, the fact that the girls are better informed about the labor market may make them invest more than boys in schooling (Betts, 1996; Jensen, 2010), and the personality of the girls may also be better aligned with the college track than the personality of the boys (Bertrand and Pan, 2013; Autor et al., 2015).

3.2 Dropout: Family background and personal characteristics

Table 1 provides the first set of regressions on how dropout from the college track relates to family background and personal characteristics. The estimated linear

---

[Benet-Martínez and John, 1998]

---

Heckman (2011) and Becker et al. (2012) show that economic preferences and psychological personality measures are complementary in explaining life outcomes and behavior.

The observed gender differences in personality are in line with what has been documented in other studies (Schmitt et al., 2008). It has also been shown that adolescence is a key period in the development of individual personality (Soto et al., 2011).
probability equations always include an indicator variable for boys, and then we add variables for family background and personal characteristics. Column (1) establishes that the observed gender difference in dropout rates is highly statistically significant, while in columns (2)-(6) we study how each of the different sets of background variables affects the estimated gender differences. In column (7), we report the regression with the full set of background variables included.

First, we note from columns (2)-(6) that adding measures of ability, preferences, and beliefs increases the estimated gender difference in dropout, while there is no effect of adding family background or measures of personality. In column (7), we observe that having a parent with some college education strongly reduces the likelihood of dropping out from the college track, which is consistent with the existing literature on intergenerational mobility in education [Bertrand et al., 2015; Bjorklund and Salvanes, 2011; Figlio et al., Forthcoming]. One mechanism that potentially explains this pattern is that parents with college background devote more time to their adolescents than less educated parents [Guryan et al., 2008], and our data provide some suggestive evidence of how parental investment in adolescents may reduce dropout rates. We observe that greater abilities, more knowledge about the labor markets, and more confidence contribute to less dropout, and the role of family background is significantly reduced when these personal characteristics are included in the regression. This is consistent with family background shaping adolescents along these dimensions. At the same time, it is interesting to note that the link between parents’ and adolescents’ education remains even when we control for a wide range of personal characteristics. Finally, we observe that the preference variables do not remain significant when all the other explanatory variables are included in the regression, and that conscientiousness is the only significant personality variable.

To shed some further light on the gender difference in dropout rates, we report in Table 2 separate regressions for boys and girls of overall dropout. We observe some interesting differences: confidence only reduces dropout among boys, while informed beliefs about the labor market and conscientiousness only reduce dropouts among girls. These differences suggest that there may be very different mechanisms at play when girls and boys drop out of school. At the same time, we observe that family background and ability are equally important for the dropout rates of boys and girls.

3.3 Dropout: Mechanisms

Dropout from the college track may happen at two stages. The adolescents may choose not to start the college track at high school and they may decide not to continue the college track after having started. Figure 3 provides an overview of the degree of dropout at each of these two levels by gender. Interestingly, we
observe that there is a striking gender difference at both stages. Girls are much more likely than boys to choose the college track (54% versus 46% percent), and girls are much less likely than boys to drop out from the college track after having started (76% versus 86%). To illustrate the importance of each of these two stages for the gender difference in overall dropout rates, we may consider what would happen if we had managed to make the boys equal to the girls at each of the two stages. If the boys chose the college track at high school at the same rate as the girls (with no change in dropout rates after having started), then the gender difference in overall dropout rates would be reduced from almost 14 percent to around 6.4 percent. On the other hand, if the boys dropped out from the college track after having started at the same rate as the girls (but with no change in the rates choosing the college track), then the gender difference in overall dropout rates would be reduced to around 8.6 percent.

In Table 2 columns (5)-(12), we study each of these two mechanisms in more detail, by reporting linear probability regressions on indicator variables of whether the adolescent chose the college track and whether he or she dropped out after having started. We observe that the choices of boys and girls are shaped by very different processes. Boys are more likely to choose the college track if they have high ability and confidence, while girls are particularly likely to choose the college track if they have informed beliefs and are conscientious. In contrast, when it comes to dropping out after having started, we observe that ability and confidence matter significantly for the girls, while only confidence matters for the boys. These differences may reflect that we are looking at a selected subsample, but they still illustrate that very different processes are at work for males and females when making the choice of whether to continue on the college track or not. Further, it is quite interesting to observe that the willingness to take risk and the willingness to compete are positively associated with both choosing the college track and dropping out after having started, particularly for the girls. This may suggest that a strong willingness to take risk or to compete may sometimes cause people to make the wrong choice, in this case the choice of the college track which may not necessarily fit the person’s abilities or interests. We observe that the same pattern does not emerge for patience. More patience increases the likelihood of girls choosing the college track, but has no effect on the likelihood of dropping out after having started. Finally, a striking finding is that while family background is highly significant for whether both boys and girls choose the college track, it does not have a statistically significant effect on whether the adolescent drops out after having started. This may suggest that family shapes the adolescents by transmitting preferences for particular educations, which provides an interesting perspective on the importance of intergenerational mobility of socioeconomic status (Roemer, 2012).
4 Concluding remarks

The findings in our study are potentially of great importance for the present debate on how to respond to the fact that boys are lagging behind in college attainment. We demonstrate that the gender difference in overall dropout rates reflects two very different mechanisms. First, boys are less likely than girls to select into the college track, and, second, boys are more likely to drop out of the college track after having started. The first mechanism may partly reflect different preferences among boys and girls, and it is not obvious that this is a difference that should call for policy intervention. The second mechanism seems more worrisome, particularly since one might imagine that the lower part of the distribution of boys choosing the college track should imply that boys who actually select into the college track are particularly motivated for the study. Family background and personal characteristics do not account for the gender difference in dropout rates after having started the college track. One possible explanation may be that the school system itself is disfavoring the boys. There is field evidence from schools (Lavy, 2008) and recent experimental evidence (Cappelen et al., 2015) suggesting gender stereotyping against boys in performance settings, which is consistent with males being more likely to drop out from high school. Our study also provides evidence that very different processes are guiding the school choices of boys and girls, which suggests that gender-specific policy interventions may be needed when aiming to reduce dropout rates. An important avenue for future research is to investigate more carefully in field settings how society can support boys and girls in adolescence, so that they make good school choices in line with their abilities and preferences.

References


Fortin, Nicole, Philip Oreopoulos, and Shelley Phipps (Forthcoming). “Leaving boys behind: Gender disparities in academic achievement,” *Journal of Human Resources*.


Figure 1: Share with college education

Note: The figure displays the historical development for the fraction of the male and female population with some college education. The fractions are based on administrative data starting in 1986.
Figure 2: Dropout from college track

Note: The figure displays the share of girls and boys that drop out from the college track. For our cohort, completing college track is defined by starting in the academic track directly after middle school and completing grade 12 by the summer of 2014.
Figure 3: Flows of participants

**Boys**

Start college track?
- Yes (54%)
  - Continue?
    - Yes (54%)
      - College track, 41%
    - No (24%)
      - Other, 13%
- No (46%)
  - Other, 46%

**Girls**

Start college track?
- Yes (64%)
  - Continue?
    - Yes (86%)
      - College track, 55%
    - No (14%)
      - Other, 9%
- No (36%)
  - Other, 36%
Figure 4: Mean differences in participant characteristics by gender

Note: The figure displays the differences in characteristics between boys and girls. Standard errors of means are indicated by the bars. “Some parental college” is a dummy indicator for whether one of the parents has some college education, the other variables are normalized to have mean zero and unit variance in the full population. For levels and standard deviations of the underlying variables, see Almås et al. (Forthcoming).
Table 1: Explaining adolescent dropout from college track

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (indicator)</td>
<td>0.133***</td>
<td>0.130***</td>
<td>0.169***</td>
<td>0.168***</td>
<td>0.164***</td>
<td>0.122***</td>
<td>0.157***</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.044)</td>
<td>(0.045)</td>
<td>(0.043)</td>
<td>(0.045)</td>
<td>(0.047)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Some parental college</td>
<td>-0.259***</td>
<td>-0.177***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.046)</td>
</tr>
<tr>
<td>Ability</td>
<td>-0.139***</td>
<td>-0.087***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.023)</td>
</tr>
<tr>
<td>Patience</td>
<td>-0.047**</td>
<td>-0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.021)</td>
</tr>
<tr>
<td>Risk-taking</td>
<td>0.024</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td>Compete</td>
<td>-0.070***</td>
<td>-0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>Beliefs labor market</td>
<td>-0.069***</td>
<td>-0.049**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td>Confidence</td>
<td>-0.124***</td>
<td>-0.058**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.025)</td>
</tr>
<tr>
<td>B5: Agreeableness</td>
<td>0.027</td>
<td>0.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.021)</td>
</tr>
<tr>
<td>B5: Conscientiousness</td>
<td>-0.104***</td>
<td>-0.066***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>B5: Extroversion</td>
<td>-0.030</td>
<td>-0.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td>B5: Neuroticism</td>
<td>0.009</td>
<td>-0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>B5: Openness</td>
<td>-0.001</td>
<td>-0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.022)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.453***</td>
<td>0.617***</td>
<td>0.435***</td>
<td>0.435***</td>
<td>0.437***</td>
<td>0.459***</td>
<td>0.551***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.042)</td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Observations</td>
<td>483</td>
<td>483</td>
<td>483</td>
<td>483</td>
<td>483</td>
<td>483</td>
<td>483</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.018</td>
<td>0.081</td>
<td>0.094</td>
<td>0.049</td>
<td>0.104</td>
<td>0.066</td>
<td>0.192</td>
</tr>
</tbody>
</table>

The table displays estimation results for linear probability models explaining an indicator for dropout from college track by participant characteristics. Robust standard errors are given in parentheses (‘$p < 0.1$, **$p < 0.05$, ***$p < 0.01$).
The table displays estimation results for linear probability models explaining dropout from college track by participant characteristics. The first four columns show results for a dummy indicating dropout as dependent variable, the middle columns show results for a dummy indicating not starting the college track as a dependent variable, whereas the four last columns show results for results for the subsample that starts on the college track using a dummy indicating dropout from this track as a dependent variable. Robust standard errors are given in parentheses (*p < 0.1, **p < 0.05, ***p < 0.01).