CESifo Working Papers
www.cesifo.org/wp

Schools, Ability, and the Socioeconomic Gradient in Education Choices

Torberg Falch
Bjarne Strøm

CESIFO WORKING PAPER NO. 3313
CATEGORY 5: ECONOMICS OF EDUCATION
JANUARY 2011

CESifo *Working Paper No. 3313*

Schools, Ability, and the Socioeconomic Gradient in Education Choices

Abstract

This paper provides evidence on the socioeconomic gradient in dropout and progression in upper secondary education in Norway. Using a rich data set covering all students transferring from compulsory education to upper secondary education in 2002, we find that student achievement at the end of compulsory school is the main predictor of dropout and delayed progression. The socioeconomic gradient is sensitive to the inclusion of prior achievement in the model. We find that the gradient is modest and non-linear, and related to dropout behavior and not grade repetition. The results are remarkably robust to controlling for a full set of school fixed effects.

JEL-Code: H420, I200, J240.

Keywords: high school dropout, socioeconomic gradient, student achievement.

Torberg Falch
Department of Economics
Norwegian University of Science and Technology
Trondheim / Norway
Torberg.Falch@svt.ntnu.no

Bjarne Strøm
Department of Economics
Norwegian University of Science and Technology
Trondheim / Norway
bjarne.strom@svt.ntnu.no

Comments from participants at the International Institute of Public Finance 2010 Congress in Uppsala are gratefully acknowledged. The data used in this paper have been collected as part of projects for the Ministry of Education and the Norwegian Directorate for Education and Training. Data is provided by Statistics Norway, and we thank Torill Vangen in SSB for her help. This paper to a large extent builds upon Byrhagen et al (2006). We thank Karen N. Byrhagen for excellent assistance on the project.

1. Introduction

A degree from upper secondary education is a prerequisite for enrolment into higher education and certification in a number of occupations. Dropout and delayed graduation influence future earnings and employment opportunities, and are potentially important inequality generating mechanisms. This is of particular importance as future demand for unskilled labor is expected to fall in most developed countries. A world wide empirical regularity is that the propensity to drop out from school at the end of compulsory education is most prevalent among individuals with limited family resources in terms of parental education and income. Knowledge of the size and nature of the socioeconomic gradient in dropout propensity is important when formulating policies to reduce income inequality. This paper studies systematically how the socioeconomic gradient in schooling decisions depends on whether prior academic ability and school effects are accounted for, using complete register data from upper secondary education in Norway.

Other studies of dropout behavior, such as Bradley and Lenton (2007) and Rice (1999) for UK, Maani and Kalb (2007) for New Zealand, and Traag and van der Velden (2008) for the Netherlands, have found that students' cognitive skills clearly reduce the probability to drop out of non-compulsory secondary education. Using US data, Belley and Lochner (2007) find that cognitive ability is a very important determinant of high school completion, although somewhat reduced in numerical terms over time, while the effect of parental income is relatively modest and stable. Further, Belley et al. (2009) show that the impact of parental income on high school completion is lower in Canada than in the US. A recent paper by Foley et al. (2009) finds that the socioeconomic gradient of high school dropout propensity in Canada is considerably reduced when controlling for cognitive ability at age 15. In particular, adding a measure of parents value placed on education effectively removes the direct impact of parental education on high school completion. Although these papers include a large number of family and student control variables, they typically include only crude measures of school characteristics, if such variables are included at all.[1]

Generally this literature has given little attention to the possibility that education motivated families may sort themselves across schools and regions in order to maximize peer group effects and obtain high teacher quality and more resources for their children. To some extent observed correlation between dropout propensity and socioeconomic background may be a result of such sorting. If, for example, the most school motivated parents tend to be located in

[1] In these respects the analysis in Maani and Kalb (2007) is representative. They include class size, the percentage of respondent's class that continues at age 16, an indicator of being in a rural area, and the local unemployment rate as regional characteristics in the model. A notable exception is Belley et al. (2009) who includes school fixed effects in their analysis of the Canadian data.

schools or areas with high teacher quality, and teacher quality differences are not properly accounted for in the empirical model, the estimated effects of family background will be biased.

Using student level register data from Norway with identifiers for both lower secondary school and upper secondary school gives a rare opportunity to handle the sorting problem in a flexible way since we can condition on a full set of fixed school effects when estimating the impact of prior academic performance and family background variables. We can estimate the impact from these variables by comparing students exposed to the same school environment at different points in their educational career. Further, the data enable us to assess the overall contribution of upper secondary school factors.

We study the determinants of deviation from expected progression using detailed information on the cohort ending their compulsory education in the spring 2002, normally the year they turn 16 year of age. Slightly above 95 percent of the cohort enrolled in upper secondary education in the fall 2002, but only about 70 percent had the expected progression at age 18 in the fall 2004, the year they should have started in their third year in upper secondary education. In the empirical analysis we relate students' deviation from expected progression to achievement at the end of compulsory schooling, a host of individual and family background variables, and school specific factors. We decompose overall deviation from expected progression into different paths; dropouts defined as students that are not enrolled in upper secondary education in fall 2004, and grade repetition defined as students that are enrolled in first or second grade at that time.

The paper is organized as follows. The next section gives an overview of institutional features of the Norwegian schooling system. Section 3 discusses theoretical and methodological issues, and Section 4 presents the data. Section 5 includes the empirical analysis, while Section 6 concludes.

2. Institutional features

The Norwegian school system consists of compulsory seven year primary and three year lower secondary education. After finishing lower secondary education the students can either choose to leave school or they can enroll into one of 15 different study tracks in upper secondary education. The latter alternative is chosen by around 95 percent of each cohort. After completing the education program in one of these tracks, students get an upper secondary school diploma qualifying for further studies (students in academic tracks) or certifying for work in a number of occupations (students in vocational tracks).

Students in a lower secondary school typically enroll in several different upper secondary schools, and upper secondary schools normally enroll students from several lower secondary schools. Most upper secondary schools offer several study tracks. The general academic study track is the largest track and includes about 40 percent of the total number of enrolled students.[2] The academic study tracks are three year programs. Vocational study tracks include industrial design, health and social work, mechanics, electrical trades, etc. They are three or four year programs and most of them include an apprentice system, either in the third and final year or in the third and fourth year, where the training is combined with commercial work in firms.

While municipalities are responsible for compulsory education, provision of upper secondary education is a county responsibility. Upper secondary education is the main service provided by the 19 counties in the country, and accounts for over 50 percent of total county spending. The counties are financed by grants from the central government. Youths with completed lower secondary education have a legal right to enroll in upper secondary education in one out of three individually ranked study tracks, a rule that is followed without exceptions by each county. Students have the right to complete upper secondary education within a period of five years after being enrolled. Students can apply for transfer to another study track after being enrolled. A transfer will delay the progression because transferred students usually have to start in the first grade in the new study track, which is included in the term "grade repetition" below.

When the number of applicants exceeds enrolment in a study track, students is ranked based strictly on the grades from lower secondary education.[3] At the end of lower secondary school, the students are given grades set by the teachers in 11 subjects on a scale from 1 (low) to 6 (high). The average grade varies slightly between subjects, from about 3.5 in mathematics to 4.3 in physical education. In addition, the students have to conduct a central exit exam in either of mathematics, Norwegian language, or English language. The overall grade used for ranking of applicants to upper secondary education is the average over all subjects, including exit exams. We use this overall average grade in the empirical analysis below.

[2] There are two additional minor academic study tracks; "Sports and physical education" and "Music, dance and drama".
[3] In addition, the algorithm takes into account that each student must be enrolled in one of the three study tracks on their priority list.

3. Theoretical issues and methodological challenges.

Theoretical issues

A natural point of departure to understand the decisions involved in upper secondary education is the standard theory on investment in human capital as originally formulated by Becker (1964) and Ben-Porath (1967). An updated discussion related to school dropout is given in Bradley and Lenton (2007). We therefore only sketch the main arguments and the implications for our empirical work.

According to the human capital model, students choose to enter upper secondary school if the expected benefits are higher than the expected costs. The benefits are represented by the expected increase in lifetime income, while costs consist of the expected foregone earnings when studying, direct costs in terms of tuition and school material, the effort required to get a diploma, and the risk of failure. As students progress through upper secondary education they receive new information. When recalculating the relevant costs and benefits, some students may find it rational to drop out of school or to move to another study track involving grade repetition. Information available might depend on parental characteristics. In addition, high prior achievement is likely related to relatively high expected benefits and low expected costs.

A particular issue is the impact of parental income on schooling decisions. Short run credit constraints are a common explanation of the observed correlation between parental income in children's adolescent years and post-compulsory school choices. Recent studies have questioned this interpretation since parental income is highly correlated over time and the relationship is substantially reduced if student ability is accounted for. Carneiro and Heckman (2002) argue that the schooling-parental income relationship most likely capture long term family effects. Belley and Lochner (2007) formulate a school choice model with schooling as a consumption good which offer another potential explanation of the correlation.

Policy-makers' typical view is that the share of students graduating from upper secondary education is too low, see for example EU (2002). This view can be rationalized by social returns to education clearly exceeding private returns or individuals making suboptimal decisions. Oreopoulos (2007) provides evidence from Canada, the US, and UK that dropouts lose an opportunity lifetime income of more than 10 percent. He argues that this evidence is best explained by myopic youth behavior involving ignoring or heavily discounting future consequences of educational decisions.

A particular concern is that dropout is most prevalent among children from families with low socioeconomic status in terms of level of education and income.[4] High ability children in families with small resources dropping out of upper secondary education have potentially a large cost for society. In addition, differences in dropout propensity by socioeconomic status have distributional consequences. Using an educational reform in the 1960's to infer causal effects, Aakvik et al. (2008) find that completing upper secondary school in Norway generates a wage premium of 10 - 15 percent. Although this evidence cannot directly be generalized to the effects in the present school system, it suggests that completing upper secondary school or not is an important source of income inequality even in Scandinavian welfare states with small wage differences. More knowledge of the effects of socioeconomic status on upper secondary education is helpful when formulating policies to deal with distributional issues.

Methodological issues

Isolating the contribution of family resources and school characteristics on educational outcomes at different stages in the educational career involve difficult methodological challenges. Cunha et al. (2010) find that early and late investments are complements in the skill formation process implying that lack of educational investment in early life is difficult to compensate in later periods. This approach requires detailed information about parental educational investments, measures of cognitive and non-cognitive ability at different ages, and a structural modeling approach.

The conventional strategy in most empirical work has been to estimate models conditioning on prior test scores, see for example Bradley and Lenton (2007), Maani and Kalb (2007) and Rice (1999). In a regression framework, this strategy gives unbiased estimates of the effect of family resources only to the extent that the error term is uncorrelated with the regressors. Todd and Wolpin (2003) give a general discussion of the required assumptions in an education production function framework.

Endogenous sorting of students and families across schools and regions is considered as an important obstacle to obtain unbiased estimates of the impact of peer group and school input variables on student performance. However, less attention has been made to the possibility that such sorting may also bias the effect of family background variables**.** First, if high ability students with school motivated parents tend to be located in schools with high teacher quality, failure to control for teacher quality differences would likely bias the effect of family resources. The same issue arises if teachers sort themselves systematically across schools according to student characteristics, as the evidence in for example Falch and Strøm (2005) suggests. Second, school motivated parents and students may sort themselves in order to

[4] See, for example, the literature review in Bradley and Lenton (2007).

maximize positive peer group effects. If peer group effects are not properly accounted for in the empirical model, this may also bias the effects of family background variables.

Including fixed school effects is an effective way of capturing sorting mechanisms.[5] In our data set we can identify for each student the lower secondary school he/she attended in addition to the upper secondary school into which he/she enrolled after finishing compulsory school. Thus, we can control for the impact of sorting across schools at different stages in the students' educational career by using both fixed lower and upper secondary school effects.

While we cannot directly claim that our approach provides unbiased estimates, we are able to investigate in a systematic way whether the effects of family background variables are robust to controlling for all characteristics of the schools. If the effects are robust, it would indicate that failure to control for endogenous sorting of families, systematic resource differences across schools, and peer effects at the school level does not represent a large problem when assessing the socioeconomic gradient in post-compulsory schooling decisions. Another strand in the literature use variance decomposition techniques to assess the role of neighborhood and family background variables on adult outcomes, see Solon et al. (2000). Raaum et al. (2006) and Lindahl (2010) use this approach on Norwegian and Swedish data, respectively, and find that neighborhoods are clearly less important than families. The evidence in Raaum et al. (2006) suggests that the impact of the neighborhood, in contrast to family background, has declined over time. While these studies use outcomes and time periods different from ours, their findings might suggest that the family background effects on upper secondary school progression are relatively stable.

4. Data description

The student data is obtained from the National Educational Database in Statistics Norway[6] and consist of all students finishing the compulsory lower secondary education in the spring 2002 and enrolling in upper secondary education in the fall 2002, including slightly above 95 percent of all graduates from lower secondary education. Thus, our sample consists of all students that made an explicit decision to enter upper secondary school.[7] The student information is matched with information of their parents, and includes school identifiers for both the lower secondary school in which they graduated and the upper secondary school in which they enrolled.

[5] Murnane et al. (1981) is an early example of using fixed school effects when estimating equations linking student achievement to prior academic performance and family background variables.
[6] http://www.ssb.no/english/mikrodata_en/datacollection/nudb/
[7] For 34 students the average grade is below one, probably because they not have completed all required courses. We drop these students from the analysis.

Table 1 presents descriptive statistics. At the start of the third year in upper secondary education, 13.5 percent of the sample is not enrolled in an upper secondary school (dropouts). 14.1 percent of the sample is enrolled in the first or second grade (grade repetition). In total, 27.6 percent of the students deviate from expected progression.

Figure 1 presents the density of the average grade from lower secondary education. Figure 1A shows that girls perform better than boys. The average grade of girls is 4.1, compared to 3.7 for boys. At the lower end of the distribution, the number of students with an average grade below three is 2.6 times higher for boys than for girls. Similarly, at the upper end, the number of students with average grade above five is 2.4 times higher for girls than for boys. Figure 1B shows that the density of the average grade for students with deviation from expected progression is clearly to the left of the density for students with the expected progression. 70 percent of the students with average grade below three have not the expected progression, while that is the case for only six percent of the students with grade of at least five.

The relationship between the average grade in lower secondary education and progression is presented in Figure 2. In the figure the students with similar grades are grouped together.[8] Figure 2A shows the strong relationship between average grade and deviation from expected progression. The share of students deviating from expected progression decreases almost linearly up to an average grade of about five. The relationships for dropouts and grade repetition are similar, although dropout is slightly more common than grade repetition for the lowest grades. Figure 2B shows that the relationship between grade and deviation from expected progression is very similar for girls and boys. If anything, for given grade, performance is slightly better for boys than for girls. Combined with Figure 1, this indicates that the larger deviation from expected progression in upper secondary school among boys than among girls is strongly related to the fact that boys have lower skills when entering upper secondary education.[9]

Figure 3 presents the density of deviation from expected progression at the school level. The dotted line shows the density across lower secondary schools, while the solid line shows the density across upper secondary schools. The variation in average lack of progression is much larger across upper secondary schools than across lower secondary schools. While it is premature to give this relationship a causal interpretation, it may indicate that school quality and peers varies more at the upper secondary level than at the lower secondary level.

[8] The first group has average grade below 1.2, the next group from 1.2 to 1.4, and so on.
[9] The average share of students with deviation from expected progression is 23 percent for girls and 32 percent for boys. The average share of dropout and grade repetition is 11 and 12 percent, respectively, for girls, and 16 percent for both outcomes for boys.

Table 1 shows that 10.7 percent of the students have at least one parent with a master or doctoral degree and 28.8 percent with a university or college degree at lower level (bachelor). More than 50 percent of the students have at least one parent with upper secondary education as their highest education, while for 5.2 percent both parents only have compulsory education. In the analysis below we include dummy variables for quartile of parental taxable labor income as measured in 2003. Table 1 shows that average parental income is five times higher in the fourth quartile than in the first quartile.[10]

96.3 percent of the sample is born in 1986,[11] and 48.1 percent enrolled in the academic track. The data include information on several other individual and parental characteristics that is not reported in Table 1 but included in the analysis below, including immigration status, parents' labor market status, and whether the mother and father live together. Our school variables include the number of students and the number of study tracks at the school. The upper secondary schools have on average about 530 students and six different study tracks. Detailed variable definitions are provided in Appendix B.

In order to analyze the impact of individual and family background variables on outcomes in upper secondary education, our starting point is to estimate variants of the following linear probability model.[12]

$$P_{ij} = \beta_1 X_i + \beta_2 G_i + \beta_3 Y_j + \varepsilon_{ij} \quad (1)$$

In the baseline specification estimated below, the dependent variable, P_{ij} is a dummy variable equal to unity if student i enrolled at school j in the fall 2002 is not in the third grade in the fall 2004, either at school j or another upper secondary school. In other words, P denotes the probability to deviate from expected progression in the beginning of the third year of upper secondary education. We will also estimate models where the dependent variable is a dummy variable for dropout or a dummy variable for grade repetition. X is a vector of individual and family characteristics, G is the average grade in lower secondary school, and Y is a vector of characteristics of the school for which the student enrolled in the fall 2002.[13] In further

[10] Average deviation from expected progression varies from 53,8 percent in the group where parents have the lowest education to 12.9 percent in the group where parents have the highest education, and it varies from 41.8 percent in the first parental income quartile to 16,1 percent in the fourth quartile.

[11] Grade repetition in lower secondary education does not take place in Norway. 2.0 percent of the sample is born in 1985, which is related to delayed school start. 0.8 percent is born before 1985. They are likely to be dropouts from lower secondary education that have returned to get the lower secondary degree. 0.9 of the sample is under-aged (born in 1987) because they started school one year earlier than normal.

[12] We have also estimated probit models, but he marginal effects for mean values of the independent variables are very close to the results from the linear model. The results are available on request.

[13] In models without lower secondary school fixed effects, we also include a set of dummy variables for the region in which the student lived in the year he was 16 years old. Statistics Norway has divided the country into 90 different labor market areas, and this is the region definition we use.

variants of the model we replace the school characteristics by upper secondary school fixed effects or lower secondary school fixed effects.

5. Empirical results

Table 2 presents the main results. The table focuses on the effect of skills, gender, parents' education and income, and characteristics of the upper secondary school in which the student enrolled the first year after finishing lower secondary school.

Baseline models

Consider first as a benchmark column (1), where the average grade in lower secondary education is excluded from the model. This specification reveals relatively large effects of individual and family background variables. The probability of deviation from expected progression is 7 percentage points higher for boys than for girls. Regarding parents education, all effects are significantly negative when compulsory education is the reference category. According to this specification, students with at least one parent with bachelor degree or higher have around 15 percentage points lower propensity to deviate from expected progression than the reference group.

To capture nonlinearities in the effect of parental income, the variable is represented by income quartiles similar to the approach in Belley and Lochner (2007) and Belley et al. (2009). The income quartile dummies shows up with significant coefficients in column (1) and suggest that students with parental income in quartile 4 are five percentage points less likely to deviate from expected progression than the reference category (quartile 1). As to the school variables, the results suggest that student's propensity to deviate from expected progression is highest in small upper secondary schools and in schools with many study tracks.

Studies from other countries suggest that prior achievement is a strong predictor of post-compulsory school dropout. The model in column (1) might thus provide a biased estimate of the socioeconomic gradient in educational choices. Column (2) in Table 2 includes students' average grade in lower secondary education in the model. Complete results for this model are presented in Appendix Table A1.[14] When average grade is included, R^2 increases from 0.17 to 0.27. A rise in the average grade of one unit, which is slightly above one standard deviation, increases the probability of normal progression with as much as 22.6 percentage points. This mirrors the relationship shown in Figure 2A. Thus, the model including both average grades and individual and family background variables more or less reproduces the raw correlation

[14] The effects of the family background variables are mainly as expected. There are strong beneficial effects of parents living together and parents' labor market participation. However, the effects of immigration status, conditional on average grade and parents' education, are weak. The dummy variables for regional labor market are jointly highly significant.

between prior achievement and the propensity to deviate from expected progression. Controlling for observed individual and family background variables does not influence the relationship between prior achievement and deviation from expected progression.

On the other hand, when achievement in lower secondary school is included in the model, the effects of all other variables drop considerably. The effect of gender does, in fact, change sign. The extended specification indicates that boys are slightly more likely than girls to follow expected progression through upper secondary school. The observation in the raw data that deviation from expected progression is higher among boys than among girls is due to the fact that boys on average have lower skills when entering upper secondary education than girls.

Regarding parents' education, the estimated effects are generally reduced by more than 50 percent when prior achievement is included in the model. The probability to deviate from expected progression is however still significantly related to parental education. Moreover, this specification reveals a highly nonlinear effect of parents' education. The effect of having at least one parent with master or PhD degree is not significantly different from zero. The nonlinear effect of parents' education is roughly consistent with evidence from other countries. Belley and Lochner (2007) and Belley et al. (2009) find that mother education above the high school level has not a significant effect on high school completion, conditional on ability. Foley et al. (2009) find that the partial effect of parental education is zero once measures of parental valuation of education are included in the model.

When conditioning on prior achievement (column 2), the estimated effect of parental income decreases by 30-50 percent compared to the results in column (1). The results suggest that, holding prior achievement constant, students with parental income in the first quartile are significantly more likely to deviate from expected progression than others. Students with parental income in the 2-4 quartiles are 2.0-2.8 percentage points less likely to deviate from expected progression. The effect size is similar to the Canadian evidence, but smaller than the US evidence. Belley et al. (2009) find that the difference in high school completion between students with parental income in fourth and first quartile is about 2 percentage points in Canada and about 11 percentage points in US, holding ability constant. While it is difficult to rigorously explain these differences, the return to education might be one relevant factor. Trostel et al. (2002) report evidence that the return to schooling is substantially lower in Canada and Norway than in the US.

We finally note that the effect of the characteristics of the upper secondary schools becomes insignificant when prior achievement is included in the model.[15]

[15] The effects of the dummy variables for labor market regions are jointly highly significant also in this model.

School fixed effects

As pointed out above, the effect of prior achievement, parental education and income, and other family background variables may be biased if students and teachers are sorted systematically across schools. In an attempt to control for such sorting, columns (3) and (4) in Table 2 includes fixed effects for upper and lower secondary schools, respectively. The coefficients in column (3) can be interpreted as an estimate for students exposed to the same school environment in upper secondary education. Thus, the model specification account for unobserved teacher quality, peer quality and other resource differences across these schools. Similarly, column (4) reports the estimated impacts for students coming from the same lower secondary school. In addition to capturing possible effects of sorting of students and teachers among lower secondary schools, this specification control for possible differences in teacher grading practices across schools. Remarkably, the inclusion of fixed school effects has a very small impact on the estimated coefficients of the included student and family background variables in the model. This result is consistent with the evidence in Belley et al. (2009) who find that inclusion of fixed high school effects does not alter the effects of family variables on high school completion in Canada. While their evidence is based on survey data, our data covers the complete population of students enrolled in upper secondary schools in Norway.

We cannot determine whether this finding reflects that characteristics of the schools are unimportant, or whether the variables of interest are uncorrelated with unobserved school characteristics. Some indication of the importance of the schools may be inferred from the change in R^2, although one should keep in mind that R^2 has non-standard properties in a linear probability model. There is a relatively small increase in R^2 when school fixed effects are included, at least compared to the change when average grade is included. Excluding all variables except the school fixed effects produces an R^2 as low as 0.05 and 0.10 when dummy variables for lower and upper secondary schools are included, respectively. If anything, this suggests that the effect of school characteristics on dropout probability and grade repetition is relatively modest.

Another way to judge the importance of schools versus other variables is to compare the distribution of the effects of individual schools in different model specifications. As a benchmark, we choose the distribution of the fixed upper secondary school effects when no other variables are included in the model. The dotted line in Figure 4 shows the benchmark distribution, while the solid line shows the distribution of the fixed school effects in the model with a full set of individual characteristics estimated in column (3) in Table 2. The figure presents densities weighted by the number of students at school, and for comparability the fixed effects are scaled to have the sample mean of the dependent variable. The figure shows that the distribution of the fixed school effects narrows considerably when the individual

characteristics are added to the model. This indicates that differences in deviation from expected progression across upper secondary schools can partly be explained by the individual student characteristics. Combined with the fact that R^2 in the model including only upper secondary school fixed effects is only 0.10, while the full model has an R^2 equal to 0.27, the results suggest that school effects on students progression is quite moderate.

Our final specification reported in column (5) in Table 2 includes fixed effects both for lower and upper secondary schools. Once more, this only produce minor changes in model coefficients. The F-values for joint significance of the fixed effects are 1.59 and 2.50 for lower and upper secondary schools, respectively, which are both highly significant.

Dropout and grade repetition

As a robustness check, we decompose our former dependent variable (deviation from expected progression) into dropouts (not registered in upper secondary education in the beginning of the third year after enrolment) and grade repetition (registered in first or second grade in the beginning of the third year after enrolment). Table 3 reports the results from the separate models for each of these outcomes. The effect of the student's average grade in lower secondary school is of similar magnitude in the two models. Regarding gender, we find that grade repetition is slightly less common for boys than for girls, given prior achievement and parental characteristics.

The effects of parents' education and income are mainly related to dropout behavior. According to the results in Table 3, the effect of parental income on the dropout probability is very similar to the effect on delayed progression in Table 2. In the model for grade repetition, the effect of parental income is very small and clearly insignificant. The results for parental education have mainly the same pattern, but with the exception of the dummy variable for master degree or PhD. This variable has a negative effect on dropout and a positive effect on grade repetition which is significant at one and five percent, respectively. Overall, however, our results clearly suggest that the socioeconomic gradient mainly appears in the dropout behavior.

6. Concluding remarks

Although most European countries and Scandinavian welfare states in particular have relatively small income differences and comprehensive education systems there is increasing concern that a significant proportion of youth fails to complete upper secondary school. This is even more important in view of the expected decrease in the relative demand for unskilled labor in the future. This paper investigates student propensity to deviate from expected progression in upper secondary school in Norway using a model that accounts for a number of

individual and family background variables along with a full set of school effects. Four main results emerge from the analysis. First, we show that the most important determinant of students' progression in non-compulsory secondary education is their achievement at the end of compulsory schooling. While it is not possible to give a clear causal interpretation of this effect, at least it suggests that the problems with non-completion of upper secondary education cannot be solved by looking at the upper secondary school system in isolation. Rather it points to the importance of educational investments made earlier in the educational career.

Second, the effects of gender and family background strongly depend on whether prior student achievement is included in the model or not. While the deviation from expected progression in upper secondary school is higher among boys than among girls in the raw data, the gender difference disappears once prior achievement is included in the model. Thus, weaker progression among boys in upper secondary school is closely related to the fact that boys enter upper secondary education with lower skills than girls.

Third, students with university educated and affluent parents are more likely to have the expected progression through upper secondary education than students whose parents only have compulsory education and are in the lower end of the income distribution. However, we find that the size of this socioeconomic gradient is relatively modest. Of particular importance is the result that the size of the socioeconomic gradient is largely reduced and highly nonlinear when prior achievement is accounted for.

Fourth, the effect of both prior achievement and family background variables is remarkably stable when we include a full set of lower secondary and upper secondary school fixed effects. Taken literally, this indicates that failure to control for endogenous sorting of families and students, systematic resource differences across schools and regions, and peer effects do not represent a large problem when assessing the socioeconomic gradient in post-compulsory schooling decisions. Whether this conclusion can be generalized to other countries with more dispersion in income distribution and living conditions than Scandinavian welfare states should be a topic for further research.

References

Aakvik, A., K. G. Salvanes and K. Vaage (2008): Measuring heterogeneity in the returns to education using an educational reform. *European Economic Review*, forthcoming

Becker, G. (1964): *Human Capital: A Theoretical Analysis of Special Reference to Education.* Columbia University Press, New York.

Ben-Porath, Y. (1967): "The production of human capital and the lifecycle of earnings". *Journal of Political Economy 75*, 352-356.

Belley, P. and L. Lochner (2007): "The changing role of family income and ability in determining educational achievement". *Journal of Human Capital 1*, 37-89.

Belley, P., M. Frenette and L. Lochner (2009): "Post-secondary attendance by parental income: Comparing the U.S. and Canada". Paper presented at the 2010 Canadian Economics Association Annual Meetings. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CEA2010&paper_id=616

Bradley, S. and p. Lenton (2007): "Dropping out of post-compulsory education in the UK: an analysis of determinants and outcomes". *Journal of Population Economics 20*, 299-328.

Byrhagen, K. N., T. Falch and B. Strøm (2006): "Frafall i videregående opplæring: Betydningen av grunnskolekarakterer, studieretning og fylke". SØF-rapport nr. 08/06, Centre for Economics Research at NTNU.

Carneiro, P. and J. Heckman (2002): "The evidence on credit constraints in post-secondary schooling". Economic Journal *112*, 705-734.

Cunha, F, J. Heckman, S. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation", *Econometrica*, 78, 883-931. .

EU (2002): "European benchmarks in education and training: Follow-up to the Lisbon European council". Communication from the commission 629.

Falch, T. and B. Strøm (2005): "Teacher turnover and non-pecuniary factors". *Economics of Education Review 24*, 611-631.

Foley, K. G. Gallipoli and D. A. Green (2009): "Ability, parental valuation of education and the high school dropout decision". Mimeo. http://grad.econ.ubc.ca/kefoley/FGG.pdf

Lindahl, L. (2010): "A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from Sweden". *Journal of Economic Inequality*, forthcoming.

Maani, S.A. and G. Kalb (2007): "Academic performance, childhood economic resources, and the choice to leave School at age 16". *Economics of Education Review 26*, 361-374.

Murnane, R., R. A. Maynard and J. C. Ohls (1981): "Home resources and children's achievement". *Review of Economics and Statistics 63*, 369-377.

Oreopoulos, P. (2007): “Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling”. *Journal of Public Economics 91*, 2213-2229.

Raaum, O., K. G. Salvanes and E. Ø. Sørensen (2006): “The neighborhood is not what it used to be”. *Economic Journal 116*, 200-222.

Rice, P. (1999): “The impact of local labour markets on investment in further education: Evidence from the England and Wales youth cohort studies”. *Journal of Population Economics 12*, 287-312.

Solon, G., M. E. Page and G. J. Duncan (2000): “Correlation between neighboring children in their subsequent educational attainment”. *Review of Economics and Statistics 82*, 383-92.

Todd, P. E. and K. I. Wolpin (2003): “On the specification and estimation of the production function for cognitive achievement”. *Economic Journal 113*, F3-33.

Traag, T and R. K. W. van der Velden (2008): ”Early school-leaving in the Netherlands. The role of student-, family- and school factors for early school-leaving in lower secondary education”. Mimeo, Research Centre for Education and the Labour Market, Maastricht University. http://www.roa.unimaas.nl/pdf_publications/2008/ROA_RM_2008_3.pdf

Trostel, P., I. Walker and P. Woolley (2002): “Estimates of the return to schooling for 28 countries”. *Labour Economics 9*, 1-16.

Table 1. Descriptive statistics

	Observations	Mean	Std. dev.
Lack of progression	52,447	0.276	.
Dropout	52,447	0.135	.
Grade repetition	52,447	0.141	
Average grade in lower secondary education	51,223	3.91	0.83
Boy	52,447	0.515	.
Parents highest education (when student 16 years old)			
Both parents lower secondary education	52,447	0.119	
At least one parent upper secondary education	52,447	0.486	
At least one parent bachelor degree	52,447	0.288	.
At least one parent master degree or PhD	52,447	0.107	.
Average parental income	52,447	285,555	198,607
Average parental income 1. quartile[a]	13,103	94,731	62,123
Average parental income 2. quartile[a]	13,112	234,831	27,863
Average parental income 3. quartile[a]	13,115	317,036	23,460
Average parental income 4. quartile[a]	13,117	495,402	261,846
Other variables			
Born in 1986	52,447	0.963	.
Academic track	52,447	0.481	.
Number of students at upper secondary school enrolled	52,251	529	274
Number of study tracks at upper secondary school enrolled	52,251	5.66	3.02

Note. [a] Measured in 2003-NOK, 1 USD ≈ 6.5 NOK.

Table 2. Determinants of deviation from expected progression. Linear probability models

	(1)	(2)	(3)	(4)	(5)
Average grade in lower secondary education	-	-0.228* (51.4)	-0.229* (53.2)	-0.236* (52.7)	-0.237* (54.3)
Boy	0.070* (14.6)	-0.014* (3.27)	-0.013* (2.89)	-0.015* (3.60)	-0.014* (3.19)
Academic track	-0.195* (25.0)	-0.057* (8.22)	-0.054* (6.59)	-0.049* (7.17)	-0.045* (5.53)
Parents highest education					
At least one parent upper secondary education	-0.070* (9.47)	-0.029* (4.28)	-0.029* (4.23)	-0.046* (4.43)	-0.044* (4.27)
At least one parent bachelor degree	-0.121* (14.5)	-0.022* (3.02)	-0.024* (3.28)	-0.039* (3.49)	-0.039* (3.45)
At least one parent master degree or PhD	-0.116* (12.0)	-0.014 (1.66)	0.007 (0.80)	0.002 (0.12)	-0.003 (0.27)
Parental income					
Parental income in quartile 2	-0.029* (4.89)	-0.020* (3.45)	-0.019* (3.33)	-0.019* (3.25)	-0.018* (3.06)
Parental income in quartile 3	-0.044* (6.77)	-0.027* (4.40)	-0.026* (4.20)	-0.027* (4.34)	-0.025* (4.09)
Parental income in quartile 4	-0.051* (6.82)	-0.026* (3.71)	-0.025* (3.53)	-0.022* (3.26)	-0.021* (3.02)
School characteristics					
Number of students at upper secondary school / 1000	-0.090* (4.07)	-0.018 (1.05)	-	-0.015 (0.82)	-
Number of study tracks at upper secondary school	0.008* (4.74)	0.002 (1.42	-	0.001 (1.00)	-
Fixed effect for upper secondary school (No of schools)	No	No	Yes (461)	No	Yes (461)
Fixed effects for lower secondary schools (No of schools)	No	No	No	Yes (1,141)	Yes (1,141)
Observations	52,251	51,156	51,156	50,833	50,833
R^2	0.154	0.258	0.276	0.284	0.300

Note. t-values adjusted for clustering at upper secondary school are reported in parentheses. * indicates significance at one percent level. In addition to reported variables, the models include dummy variables for year of birth, immigration status, civil status of father and mother, labor market status of father and mother. Dummies for labor market region are included except in models with fixed lower secondary school effects. Full results for the models in columns (2) and (5) are reported in Appendix Table A1.

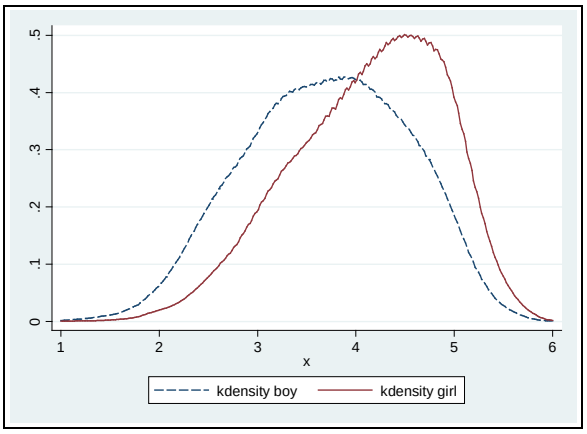
Table 3. Determinants of dropout and grade repetition. Linear probability models

	(1)	(2)	(3)	(4)
Dependent variable	Dropout		Grade repetition	
Average grade in lower secondary education	-0.106* (29.2)	-0.110* (29.4)	-0.122* (40.6)	-0.127* (40.5)
Boy	-0.001 (0.15)	-0.004 (1.10)	-0.014* (4.19)	-0.010* (3.14)
Academic track	-0.066* (13.4)	-0.068* (11.2)	0.009 (1.74)	0.022* (3.49)
Parents highest education				
At least one parent upper secondary education	-0.052* (4.96)	-0.048* (4.49)	-0.0002 (0.02)	0.004 (0.38)
At least one parent bachelor degree	-0.050* (4.49)	-0.047* (4.15)	0.003 (0.32)	0.009 (0.85)
At least one parent master degree or PhD	-0.032* (2.76)	-0.032* (2.76)	0.022 (2.04)	0.029* (2.64)
Parental income				
Parental income in quartile 2	-0.016* (3.06)	-0.014* (2.75)	-0.004 (0.80)	-0.004 (0.67)
Parental income in quartile 3	-0.024* (4.56)	-0.022* (4.21)	-0.003 (0.63)	-0.003 (0.55)
Parental income in quartile 4	-0.020* (3.47)	-0.017* (2.99)	-0.007 (1.07)	-0.003 (0.55)
School characteristics				
Number of students at upper secondary school * 1000	-0.007 (0.56)	-	-0.010 (0.97)	-
Number of study tracks at upper secondary school	0.0003 (0.56)	-	0.002 (1.49)	-
Fixed effects for lower secondary schools (No of schools)	No	Yes (461)	No	Yes (461)
Fixed effect for upper secondary school (No of schools)	No	Yes (1,141)	No	Yes (1,141)
Observations	51,156	50,833	51,156	50,833
R^2	0.127	0.172	0.103	0.146

Note. t-values adjusted for clustering at upper secondary school are reported in parentheses. * indicates significance at one percent level. In addition to reported variables, the models include dummy variables for year of birth, immigration status, civil status of father and mother, labor market status of father and mother. Dummies for labor market region are included except in models with fixed lower secondary school effects. Full results for the models in columns (2) and (4) are reported in Appendix Table A1.

Figure 1. The distribution of average grades in lower secondary education

A. Girls and boys



B. Progression

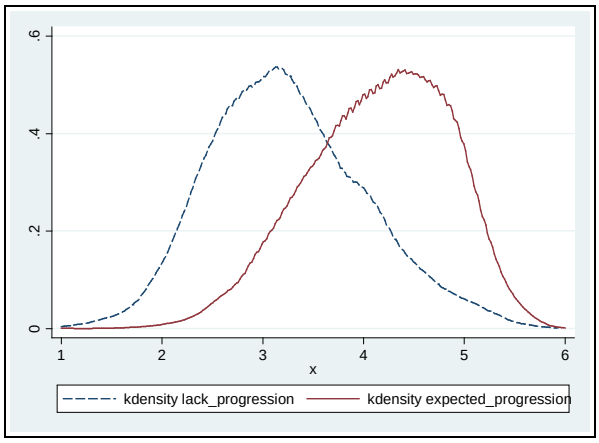
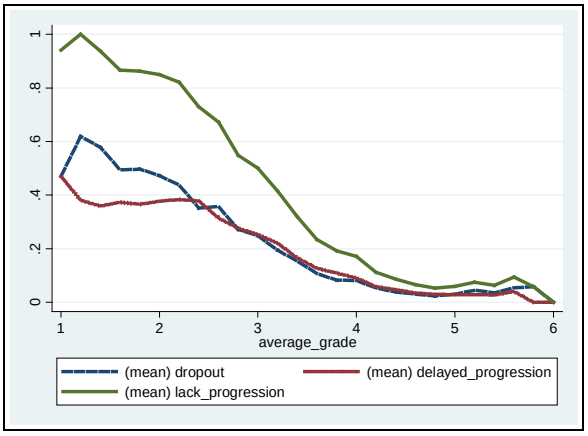


Figure 2. Average grade and lack of progression

A. Type of lack of progression



B. Gender differences

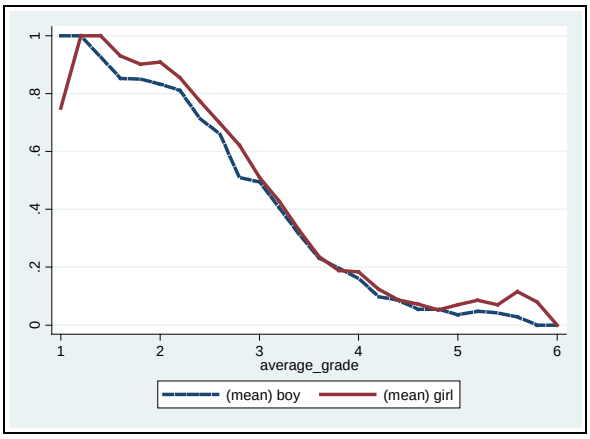


Figure 3. Density of deviation from expected progression at the school level

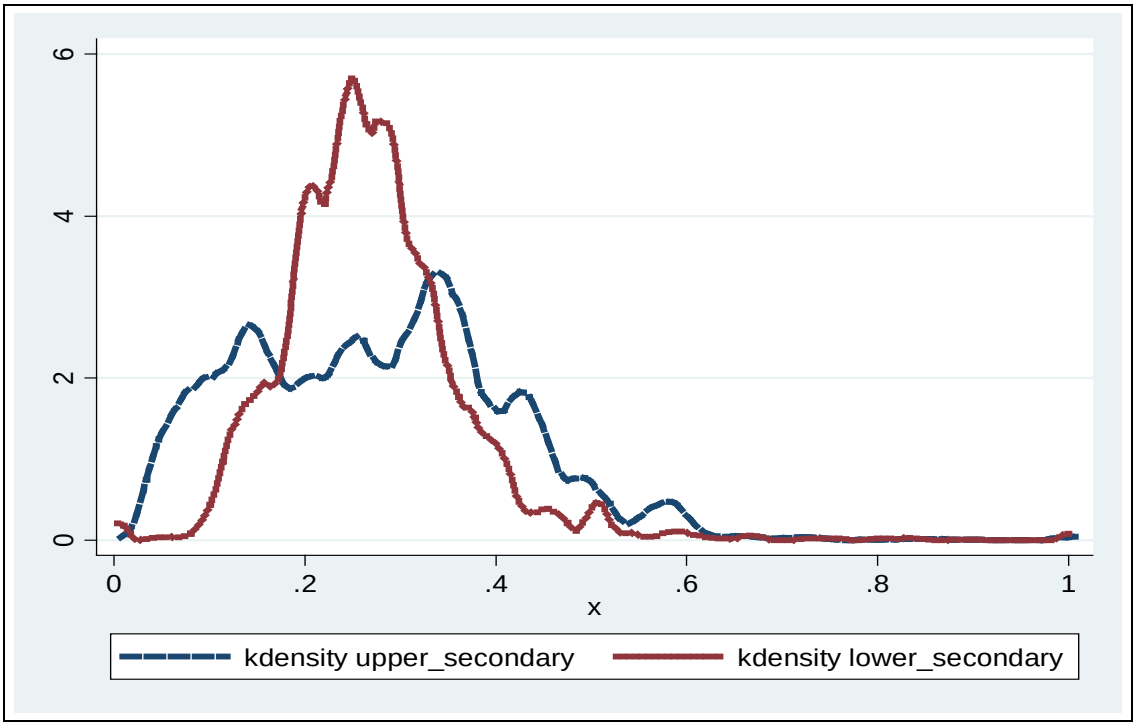
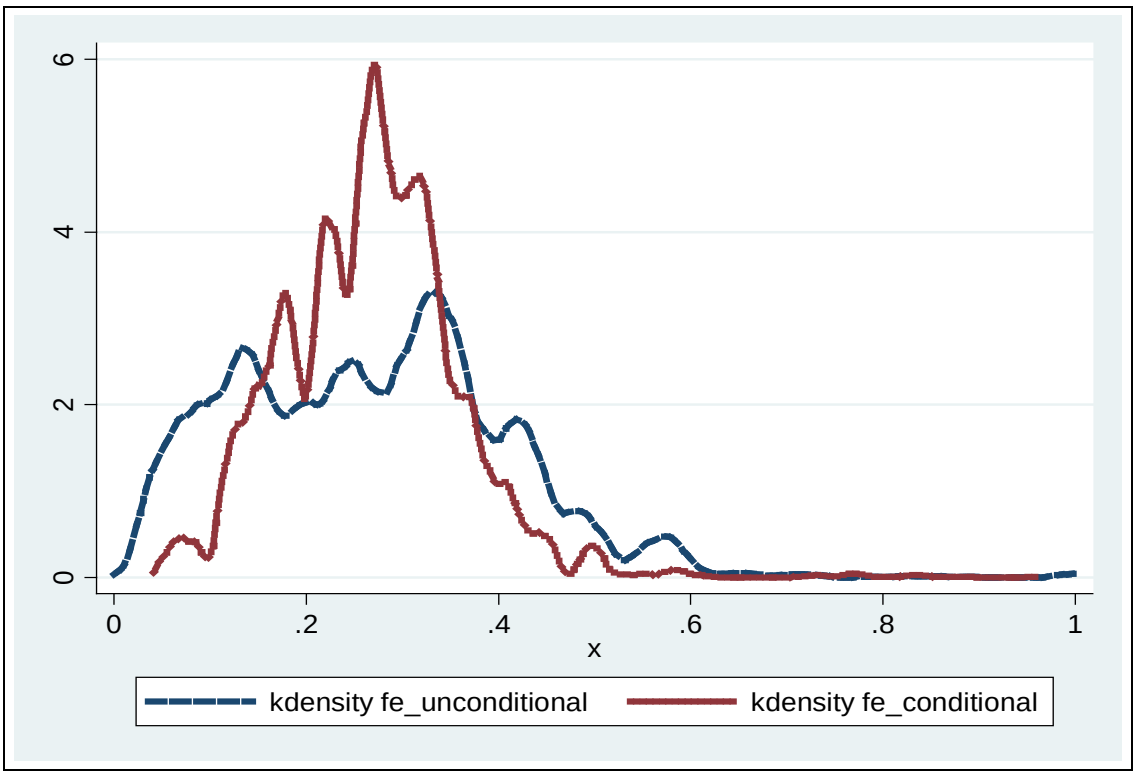


Figure 4. Density of conditional and unconditional fixed effects for upper secondary schools



Appendix Table A1. Complete results

	(1)	(2)	(3)	(4)
Dependent variable	Deviation from expected progression		Dropout	Grade repetition
Average grade in lower secondary education	-0.228* (51.6)	-0.237* (54.3)	-0.110* (29.4)	-0.127* (40.5)
Boy	-0.014* (-3.26)	-0.014* (-3.19)	-0.004 (-1.10)	-0.010* (3.14)
Academic track	-0.057* (8.27)	-0.046* (5.53)	-0.068* (11.2)	0.022* (3.49)
Parents highest education				
At least one parent upper secondary education	-0.052* (5.07)	-0.044* (4.27)	-0.048* (4.49)	0.004 (0.38)
At least one parent bachelor degree	-0.047* (4.22)	-0.039* (3.45)	-0.047* (4.15)	0.009 (0.85)
At least one parent master degree or PhD	-0.010 (0.82)	-0.003 (0.27)	-0.032* (2.76)	0.029* (2.64)
Parental income				
Parental income in quartile 2	-0.020* (3.49)	-0.018* (3.06)	-0.014* (2.75)	-0.004 (0.67)
Parental income in quartile 3	-0.028* (4.50)	-0.025* (4.09)	-0.022* (4.21)	-0.003 (0.55)
Parental income in quartile 4	-0.027* (3.82)	-0.021* (3.02)	-0.017* (2.99)	-0.003 (0.55)
Immigration status				
Immigrant from western country, first generation	0.036 (1.32)	0.033 (1.20)	0.034 (1.37)	-0.001 (0.06)
Immigrant from western country, second generation	0.052 (1.06)	0.054 (1.09)	0.037 (0.95)	0.017 (0.46)
Immigrant from non-western country, first generation	-0.029* (2.77)	-0.029* (2.69)	-0.034* (3.50)	0.004 (0.43)
Immigrant from non-western country, second generation	-0.025 (1.58)	-0.020 (1.22)	-0.041* (3.15)	0.021 (1.73)
Family structure				
Live with mother and not with father	0.060* (11.3)	0.059* (11.2)	0.023* (5.32)	0.035* (7.82)
Live with father and not with mother	0.043* (4.85)	0.042* (4.75)	0.016 (2.23)	0.026* (3.49)
Does not live with either father or mother	0.073* (10.5)	0.069* (9.77)	0.027* (4.99)	0.042* (6.60)
Parents employment status				
Mother working full time	-0.035* (6.66)	-0.031* (5.96)	-0.016* (3.57)	-0.015* (3.53)
Mother working part time	-0.036* (6.22)	-0.035* (5.95)	-0.021* (4.42)	-0.014* (3.08)
Mother self-employed	-0.024* (2.59)	-0.024* (2.67)	-0.010 (1.32)	-0.014 (1.82)

	(1)	(2)	(3)	(4)
Father working full time	-0.023* (3.75)	-0.018* (3.13)	-0.007 (1.37)	-0.011 (2.08)
Father working part time	-0.005 (0.47)	-0.0003 (0.03)	0.006 (0.72)	-0.006 (0.70)
Father self-employed	-0.028* (3.60)	-0.028* (3.68)	-0.013 (1.88)	-0.015 (2.25)
Year of birth				
Born before 1981	-0.031 (0.40)	-0.090 (0.53)	0.074 (0.37)	-0.164 (0.85)
Born in 1981	0.095 (0.77)	-0.229 (0.95)	-0.279* (5.84)	0.050 (0.22)
Born in 1982	-0.054 (0.63)	-0.347* (2.59)	-0.078 (0.62)	-0.269* (3.29)
Born in 1983	-0.053 (0.78)	-0.160 (1.20)	-0.025 (0.19)	-0.135 (1.22)
Born in 1984	0.104 (1.89)	0.087 (0.99)	-0.083 (1.18)	0.170 (1.84)
Born in 1985	0.093* (6.26)	0.095* (6.46)	0.057* (4.09)	0.038* (2.72)
Born in 1987	0.057* (3.17)	0.036 (2.07)	0.023 (1.87)	0.013 (0.96)
School characteristics				
Number of students at upper secondary school * 1000	-0.018 (1.05)	-	-	-
Number of study tracks at upper secondary school	0.002 (1.42)	-	-	-
Regional labor market FE, F(89, ≈460)	14.1	-	-	-
Fixed effects for lower secondary schools (No. of schools)	No	Yes (461)	Yes (461)	Yes (461)
Fixed effect for upper secondary school (No. of schools)	No	Yes (1,141)	Yes (1,141)	Yes (1,141)
Observations	51,156	50,833	50,833	50,833
R^2	0.258	0.300	0.172	0.146

Note. t-values adjusted for clustering at upper secondary school are reported in parentheses. * indicates significance at one percent level.

Appendix B: Variable definitions.

Dependent variables:

Deviation from expected progression: Binary dependent variable equal 1 if individual was not registered or was registered in 1 or 2. grade in fall 2004, 0 if registered in 3.grade in fall 2004.

Dropout: Binary dependent variable equal to 1 if not registered in upper secondary school in fall 2004.
Grade repetition: Binary dependent variable equal to 1 if registered in 1. or 2.grade in fall 2004, 0 if not registered or registered in grade 3.

Explanatory variables:

Boy: Binary explanatory variable equal to 1 if student is a boy, 0 if girl

Academic track: Binary explanatory variable equal to 1 if enrolled in one of the three study tracks that give access to higher education, 0 otherwise.

Both parents compulsory education: binary variable, registered in the year the student was16 years old.

At least one parent upper secondary education, binary variable, registered in the year the student was16 years old.

At least one parent bachelor degree: Binary variable, registered in the year the student was16 years old.

At leat one parent master or Phd degree: Binary variable, registered in the year the student was16 years old.

Parental income in quartile i: Binary variable equal 1 if sum of father's and mother's taxable income from work in 2003 is in the i'th.quartile, 0 otherwise.

Immigrant from western country, first generation: Binary variable equal 1 if student born abroad with both parents born in western country outside Norway, 0 otherwise.

Immigrant from non-western country, first generation: Binary variable equal 1 if student born abroad with both parents born in non-western country, 0 otherwise

Immigrant from western country, second generation: Binary variable equal 1 if student born in Norway, with both both parents born in western country outside Norway, 0 otherwise.

Immigrant from non-western country, second generation: Binary variable equal 1 if student born in Norway, with both parents born in non-western country, 0 otherwise.

Live with mother and not with father: Binary variable equal 1 if student lives with biological mother and not with father in the year the student was16 years old, 0 otherwise.

Live with father and not with mother: Binary variable equal 1 if student lives with biological father, and not with mother in the year the student was16 years old, 0 otherwise.

Does not live with either mother or father: Binary variable equal 1 if student lives with neither biological mother nor biological father in the year the student was16 years old, 0 otherwise.

Mother (Father) working full time 2004: Binary variable equal 1 if student's mother (father) working as employee more than or equal to 30 hours a week, 0 otherwise

Mother (Father) working part time 2004: Binary variable equal 1 if student's mother (father) working as employee less than 30 hours a week, 0 otherwise

Mother (Father) selfemployed 2004: Binary variable equal 1 if student's mother (father) selfemployed.