Das Sozio-
oekonomische
Panel

713

2014

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP — The German Socio-Economic Panel Study at DIW Berlin

713-2014

Who Dares, Wins?
A sibling analysis of tertiary education transition in Germany

Tamás Keller and Guido Neidhöfer

DIW SOEP

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Who Dares, Wins?
A sibling analysis of tertiary education transition in Germany*

Tamás Keller[ab] and Guido Neidhöfer[c]

[a]) TÁRKI Social Research Institute Budapest; [b]) Research Center for Educational and Network Studies, Hungarian Academy of Science; [c]) School of Business & Economics, Freie Universität Berlin

26. November, 2014

Abstract. Past studies have found that parental background has a considerable impact on educational decisions. Our knowledge is, however, still limited regarding educational transitions later in life, such as into tertiary education. Is parental background a predominant factor in this relatively late educational decision, or do individual talent and determination have an impact of their own? We address this question by decomposing the probability of success – regarded by rational choice models in sociology as a major component in the explanation of educational choices – as a function of observable and unobservable characteristics, using school grades and subjective perceptions about future educational success. To control for the overall effect of family background, a sibling analysis is performed. The data is derived from the German Socioeconomic Panel (SOEP), where we can follow those pupils who participated in the survey at the age of 17 later in life. Our results are twofold. Parental background (through school grades) exerts a strong influence at the time of transition to university; however, subjective perceptions also have an effect that is independent of parental background.

JEL Codes: I23, I24, J62
Keywords: Tertiary education transition; Sibling analysis; Subjective perceptions; Rational choice theory; Equality of educational opportunities

* We would like to extend our cordial thanks to Timm Bönke for the help provided in the construction of the data set; Miriam Gensowski, Robin Jessen and Maximilian Stockhausen for insightful comments; Vanessa Desche for her exhaustive research assistance; and to those who participated in seminars and conferences in Berlin, Budapest, Kassel, Potsdam and Vienna for their useful remarks. This project has been supported by grants from the Alexander von Humboldt Foundation and from the OTKA (Hungarian Scientific Research Fund), grant number: PD 105976. Guido Neidhöfer thanks the Hans Böckler Foundation for its financial support. Any errors are entirely ours.

Corresponding author: Tamás Keller, TÁRKI Social Research Institute and Research Center for Educational and Network Studies, Hungarian Academy of Science, former Alexander von Humboldt Research Fellow at Freie Universität Berlin, School of Business and Economics. Address: Budaörsi út 45, 1118 Budapest, Hungary (keller@tarki.hu).

I. Introduction

Social mobility is one of the key issues addressed in debates about equality of opportunity in society. Indeed, taking an intergenerational context, if the goal is to *level the playing field*, probably no one would argue in favour of a high incidence of economic status being passed down from parents to children (Roemer, 2000). However, it is also well known that the underlying mechanism of intergenerational transmission is driven by different forces which, in very simple terms, can be summarized as the effects of nature and nurture.[1] All these aspects must be taken into account if our aim is to evaluate the net contribution of individual motivations or efforts.

An important issue here is to investigate the role of both individual cognitive and non-cognitive skills. From the findings of previous research, we know that the two groups of skills play an equally important role in schooling decisions and other economic outcomes (Heckman, Stixrud and Urzua, 2006). However, recent analysis shows that school performance is not equally distributed among pupils with different parental background: pupils in higher social classes generally perform better at school (OECD, 2010). Furthermore, the impact of parental background remains stable even controlling for ability – which is passed from parents to children through genetic, psychological and cultural channels (referred to in sociology as *primary effects*); this shows that parental background has a (*secondary*) effect, which is regarded as being directly connected to educational choices (Boudon, 1974).

As previous literature shows, different educational choices may also be influenced by social class differences in preference ordering (Gambetta, 1987), goal setting (Keller and Zavalloni 1964), educational demand (Murphy, 1981, 1990), and risk aversion and expectations of success (Breen and Goldthorpe, 1997). Former empirical analysis has also shown that, even controlling for ability, non-cognitive skills such as risk aversion (Hartlaub and Schneider, 2012), interpretation of success (Stocké, 2007) and subjectively perceived success probabilities (Tolsma, Need and Jong, 2010) play a crucial role in educational decisions. But interestingly enough, the intergenerational transmission effect in the case of non-cognitive skills seems to be considerably lower than in the case of cognitive skills (Anger, 2012). Thinking about policy decisions, this highlights the opportunity to focus more on non-cognitive skills, since these seem to be less determined by parental background than cognitive skills, though their impact on economic outcomes may be at least as great. But so far our knowledge has been limited in this field by issues related to data availability.

In the educational literature much attention has been devoted to school effects or unobserved school heterogeneity (e.g. Falch and Strøm, 2013), and less to unobserved family heterogeneity and its influence on non-cognitive abilities. Previous research has usually focused only on individual aspects of parental background – such as occupational, educational and cultural status or income – but has not controlled for the overall impact of family characteristics. This is quite surprising, since (as previous research has shown) the impact of family heterogeneity in unobserved characteristics seems to be far more influential in educational choice than such widely credited explanatory variables as parental income (Tamm, 2008).

One important aspect of the role of family background in this mechanism could be the transmission from parents to children of values that are essential in occupational or educational choices (Corneo and Jeanne, 2010). So an important question is whether the individual component is still decisive in educational decisions. Here, previous studies have shown that by focusing on parent–child similarities, only one part of family effects is captured (as discussed by, for example, Schnitzlein, 2013). Analysing the family impact through

[1] Recent insights into this topic have shown that both genes and environment contribute significantly to the intergenerational transmission of income and education. For a review of the literature on this subject, see Sacerdote (2011).

sibling correlations seems, therefore, to be more appropriate (Björklund and Jäntti, 2012). Controlling for observable and unobservable similarities between siblings provides an opportunity to control for the full range of family heterogeneity.

But we need a framework to understand the determinants of differing educational decisions, taking account of subjective perceptions. Rational choice models in sociology regard perceived probability of success as a major component in the explanation of educational choices (Breen and Goldthorpe, 1997; Erikson and Jonsson, 1996; Esser, 1999). So, following on from this theoretical basis, our aim is to estimate the impact of perceived probability of success on applications to enter tertiary education in Germany. Here we face two major challenges: first, estimates of success probability might not be exclusively derived from previous school performance; subjective parameters could also play an important role. Second, unobserved family-related factors might shape these subjective parameters strongly. Previous analysis that has observed subjective estimations has usually focused on the role of parental (not adolescent) aspirations (Paulus and Blossfeld, 2007) or satisfaction (Pietsch and Stubbe, 2007). This was probably inspired by previous empirical findings that identified the ways in which parents take care of their children's education (checking their homework, enforcing rules at home, spending time together, etc.) and that identified significant effects of such behaviours on children's school performance (Baker and Stevenson, 1986). But undoubtedly, paucity of data is also an issue that has impeded more profound analysis.

To tackle this issue, we construct our sample from a household survey database, where we can account for the school grades and personal perceptions of future success among pupils at the age of 17, and can follow their actual educational decisions later in life. This gives us a unique opportunity to extend the analysis, differentiating first between objective and subjective measurements (i.e. school grades and personal perceptions), and then controlling for family heterogeneity, using characteristics shared by siblings. The exercise is worthwhile, since it allows us to address two questions of a different nature. The first, of general character, is whether subjectively perceived success has an independent impact on educational choices, or whether it simply arises out of the transmission mechanism within families. The second, considering the widely studied high selectivity of the German secondary education system, deals with the equality of educational opportunities at the time of transition to tertiary education.

This paper is organized as follows. Section II deals with selectivity at transition to tertiary education and the German education system. Section III summarizes some insights from rational choice theory and discusses some theoretical considerations in terms of the role of subjective perceptions in educational decisions. Section IV presents the estimation strategy, while in Sections V and VI we describe the data and offer some first, descriptive evidence. The main results are presented and discussed in Section VII. Sections VIII and IX consider further the possible causes of endogeneity and report on the robustness checks performed. Finally, Section X contains our concluding remarks.

II. Selection in the school system and choice of going on to tertiary education in Germany

In Germany, after primary school the school system is divided into three tracks (typically lasting four years; or six years in two of the country's federal states) of which only the highest – upper secondary school (*Gymnasium*) – leads directly to university. The aim of the other two lower secondary education tracks is to give some general education and prepare pupils for manual or technical work.[2] This practice of sorting pupils very early on in their educational career could be one of the main causes of the strong influence of parental socioeconomic background that has been experienced in Germany over several decades (Heineck and Riphahn, 2009) and that stands in contrast to the situation in other countries (Woessmann, 2008), as several studies have shown (e.g. Bauer and Riphahn, 2006; Brunello and Checchi, 2007; Dustmann, 2004; Hanushek and Woessmann, 2006).

Sociological research has come up with similar findings on this point: equality of educational opportunity in Germany is fundamentally challenged by early streaming, and social origin seems to have a definitely smaller effect on the educational career at older ages (Schneider, 2008).[3] It has also been underlined that the consequences of unchallenging educational decisions – namely *underachievement* – cannot entirely be explained by status differences in terms of academic advantage or personality traits (Uhlig, Solga and Schupp, 2009). As a paradox of the German educational system, lower-status pupils face a dual barrier on the road to the highest track (*Gymnasium*): they need greater achievements in order to get the same school-track recommendation from their primary-school teachers; and they have to 'convince' their parents by their higher achievements that they should be launched on the academic trajectory (Pietsch and Stubbe, 2007).

That said, an interesting question is what happens with those pupils who are selected for *Gymnasium*, make it onto the highest educational track and achieve their final school diploma, which allows them to go on to university. Riphahn and Schieferdecker (2010) show that the effect of parental background on the transition to tertiary education in Germany is strong (though it has been declining over time), especially in terms of parental income. This result holds even controlling for the selectivity in secondary school. Other studies of the transition to tertiary education in Germany report that parental social class has an impact, especially via secondary effects which are not directly related to school performance (Neugebauer and Schindler, 2012; Schindler and Reimer, 2010). However, previous analysis has taken account neither of subjective parameters nor of unobserved family heterogeneity.

III. The importance of perceived probability of success – some theoretical considerations

Since the educational expansion of the late 60s and early 70s, educational decisions have been at the focus of research into inequality of opportunity (see, for example, Bratti et al., 2008). Following standard human capital theory, educational choices have been seen primarily as a function of the expected returns to education – discounted by the time that is needed to accumulate a certain amount of human capital (or, in other words, to achieve a certain educational level) – and opportunity costs. However, by measuring these in monetary terms – e.g. (expected) future earnings – one might lose important aspects that shape the educational decisions of individuals from different families and social classes (Akerlof, 1997). Nor does the standard framework take into account what might influence the formation of different expectations among different people. For example, subjective perceptions about educational success should strongly influence expectations about future labour market outcomes.

One of the existing theoretical frameworks for understand the dynamics behind this phenomenon and the role of subjective perceptions is rational choice theory in sociology (summarized by Stocké, 2010). The differences between particular approaches (Breen and

[2] Apart from the three educational tracks described here, there are other, less common tracks, such as an integrated school type (*Gesamtschule*). Since the *Gesamtschule* may also lead to the leaving diploma (*Abitur*), we include pupils of this school type. The analysis of the others goes beyond the scope of this work.

[3] An interesting survey on the definition of equality of educational opportunity, including an application to higher education access, can be found in Brunori et al. (2012).

Goldthorpe, 1997; Erikson and Jonsson, 1996; Esser, 1999) are minimal, and basically lie in the definition of benefits offered by a particular educational option. Erikson and Jonsson (1996), for example, consider the utility of choosing a particular educational option as the product of benefits (*B*) and perceived success (*p*), minus the cost (*C*) of education, hence $U = p \times B - C$. Cost and benefits are measured in the same units, and not exclusively in monetary terms but also in psychological categories. Furthermore, higher social classes attach a higher value to the same level of education. This is because maintaining parental social class position is an important driver in educational decisions, and offspring in higher classes need to attain relatively higher educational levels to avoid downward mobility.[4] However, since the utility to choose a particular educational option is the product of benefits and probability of success, a high level of success probability could compensate for a low benefit level and therefore increase the utility of a particular educational option. This could be important among low-status pupils who do not attach a high value to a high level of education.

Similar implications arise by using other approaches, such as the model of educational choices proposed by Esser (1999). He argues that the choice of a particular educational level occurs if educational motivations are greater than investment risks. Thus, on the left-hand side of this inequality is educational motivation, defined as benefits arising from a particular educational option and the cost of the status loss occurring if a certain educational level is not achieved. On the right-hand side is investment risk, interpreted as the cost of education divided by the probability of success. Since investment risk diminishes the higher the probability of success is perceived to be, a high probability of success could level out the small educational motivation in the above-mentioned inequality and thus increase the likelihood of opting for a more competitive educational scenario.

The theories described so far offer some possible explanations for the complex mechanism underlying different educational choices and the role of perceived probability of success. The role of the family in the general framework is treated as an intrinsic factor, assuming that choices and decisions arise from the parents' socioeconomic background. To create a direct link to the inheritability of wealth, income or social status, we could generalize, recalling the seminal work of Becker and Tomes (1979), and refer to these as *family 'endowments' transferred from parents to children*. In the sociological literature, for example, it has been argued that higher-status classes have higher aspirations (Rosen, 1956) and are motivated by different values (Gambetta, 1987). In the latter work, this mechanism is called *over-adaption*, and is interpreted as a *pushing* factor, which constrains individual opportunities. Gambetta's framework suggests thus that characteristics which are independent of parental background or other factors *pushing* educational choices might act as a *jumping point* out of the deterministic cycle.

According to this line, our contribution is the empirical analysis of perceived probability of success for the educational choice after secondary education. Here, we deal with probability of success not only as determined by ability or school performance (which is usually measured by school grades), but also as influenced by subjective parameters. We focus especially on the individual component, of these two effects (school performance and subjective estimations about future educational outlooks) – namely the parts that are independent of given circumstances, like family background.

IV. Empirical strategy

The aim of our analysis is to decompose the factors that influence educational choices after secondary education, isolating parental background effects and determining the part that is

[4] For an empirical analysis of testable hypotheses extracted from rational choice models, see Davies, Heinesen and Holm (2002)

due to individual characteristics. We start by evaluating the objective and the subjective dimensions of ability, as measured by school grades and subjectively estimated future success, respectively.

The unobserved family-level heterogeneity, however, could correlate with the regressors, and therefore parameter estimations could be biased. Basically, two approaches have been taken in the literature: random effects models (which use 'within' and 'between' variance) and fixed effects models (which use 'within' variance). The aim of random effects models is to measure the proportion of variance that can be attributed – in this case – to parental background. When the sample is restricted to clusters with at least two individuals (i.e. families with two siblings), the shared variance can be interpreted as sibling correlation. Following Solon, Corcoran, Gordon and Laren (1991) the sibling correlation can be rendered as

$$\rho = corr(s_{1j}, s_{2j}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}$$

where s_{1j} and s_{2j} are the outcomes of two different siblings in family j. The *sigmas* derive from a variance decomposition of the error term ε in an equation $s_{ij} = \beta' X_{ij} + \varepsilon_{ij}$, which is subdivided into a permanent component shared by siblings (a) and another component not shared by siblings in the same family (u); formally $\sigma_\varepsilon^2 = \sigma_a^2 + \sigma_u^2$. This model assumes that individual and family components are orthogonal to each other.

By definition, random effects models assume that there is no correlation between the independent variables and the group heterogeneity. If this is a plausible assumption, the random effects model is more effective than a fixed effect model (Wooldridge, 2009: 493). As a way of relaxing this assumption, Mundlak (1978) proposed adding group means of independent variables (which vary within groups) to the regressors. If the group means are jointly not significant, the hypothesis that individuals are randomly assigned to groups cannot be rejected. One advantage of using random effects models is that we can distinguish between family and individual effects, including the average value among siblings (family-level variance) and the individual deviation (individual-level variance) from it as independent variables. We apply a so-called 'random intercept model', and in this particular case we are interested in both individual and family effects.

Taking the above into account, we construct a simple econometric model displaying educational choices. Thus, let uni be a variable which is 1 if the individual goes to university, and 0 otherwise. Then, the latent variable $uni*$ displays the unobserved utility of individual i belonging to family j to choose university. It can be modelled as a function of certain individual and family-specific characteristics of interest (a detailed definition of the variables can be found in Table 1):

$$uni_{ij}^* = f\{\alpha_0(grade_{ij}) + \beta_0(success_{ij}) + \gamma_0(status_j) + \Psi_j\varepsilon' + \Pi_{ij}\delta' + a_j + u_{ij}\} \quad (1)$$

Here, $grade$ represents the school grades, $success$ the perceived probability of success, and $status$ parental social status, while Ψ and Π contain individual and family-specific control variables. (All these variables are described in detail in Section V.) Equation (1) is first estimated by simple Ordinary Least Squares (OLS) (Model 1), Logit (Model 2) and then by more sophisticated random effects models (Models 3 and 4), where we control for the overall family effect due to unobservable characteristics.

Next, we try to measure unobservable effects through sibling averages of the relevant variables, in order to see if individual variations from this average still have a significant impact on the choice of tertiary education. This technique is often applied in multilevel analysis, such as in educational studies using school or class averages, and has the advantage that individual deviations from the cluster mean are, by construction, independent of the unobserved cluster heterogeneity (see e.g. Raudenbush and Bryk, 1986; for a more general approach, see Raudenbush and Bryk, 2002). The following equation represents the relationship mentioned here:

$$uni_{ij}^{*} = f\{\alpha_1(\Delta grade_{ij}) + \alpha_2(\overline{grade}_j) + \beta_1(\Delta success_{ij}) + \beta_2(\overline{success}_j) + \gamma_0(status_j) + \Psi_j\varepsilon' + \Pi_{ij}\delta' + a_j + u_{ij}\} \quad (2)$$

where $\overline{grade}$ and $\overline{success}$ are the average across all siblings in the same family, and Δ denotes the individual deviation from this average for the two variables. In contrast to equation (1), equation (2) is estimated first only for individuals with at least one sibling for whom data is available (though the sibling did not necessarily sit the *Abitur*) (Model 5), and then only for those whose siblings all obtained the *Abitur* (Model 6).

We call Model 1 and Model 2 *Basic Models*. These models do not have any specific condition about the family effect. Models 3 and 4 are called *Family Models*, since they assume specific conditions about family effects, but are less effective at estimating the shared variance among siblings, since some individuals in the sample have no siblings. To overcome these shortcomings, the *Sibling Models* (Models 5 and 6) are employed. From the estimation of Model 6, we obtain the sibling correlation explained at the beginning of this section. Standard errors are calculated using the delta method (Oehlert, 1992) and clustering among siblings.

V. Data and variables

The data we use is derived from the German Socioeconomic Panel 2012 (SOEP v29). SOEP is a representative household panel survey which is ideal for our research purposes, thanks to its wealth of information on both individual characteristics and family background (for further information on the German Socioeconomic Panel, see Wagner, Frick and Schupp, 2007). Young people living in a panel household are interviewed when they reach the age of 17, using a special Youth Questionnaire in SOEP. Since most of them are interviewed again in the following years, this feature provides a unique opportunity to connect their answers at this stage in their lives with later outcomes.[5]

Actually, SOEP is not particularly designed for educational research. However, in a special data set on educational participation and transition, efforts have been made to collect useful information about individual educational careers, based on some plausible assumptions (see Lohmann and Witzke, 2011). For the purposes of our study, we could indirectly identify those who obtained the highest secondary school certificate (*Abitur*), using information about the 'first exit' from secondary school (i.e. when someone is observed to be in secondary education at one given wave of SOEP, but not at the next). If 'first exit' occurred before the age of 17, the person concerned was regarded as a school dropout.

The dependent variable in the analysis is categorical in nature (uni). It is coded 1 if someone went to university (or a university of applied science)[6] after *Abitur*, and 0 otherwise. Here, only the first (educational) choice after *Abitur* is considered, and we do not set any limit on when the entry to tertiary education occurs.[7] So since the observed time period is restricted

[5] For a detailed explanation of the Youth Questionnaire (BIOAGE17), see SOEP v29 Documentation (SOEP-Group, 2013) pp. 138ff.

[6] Since we also included universities of applied science (*Fachhochschulen*), and since it is also possible to attend these with a lower type of secondary school diploma (*Fachabitur*), we also included pupils with *Fachabitur* in the notion of those who made *Abitur*. An interactive overview of the German educational system, showing the possible paths of transition from various kinds of school types, can be found at: http://www.bpb.de/fsd/bildungsgrafik2

[7] In our sample, pupils were around 19 years of age when they obtained their *Abitur* (the youngest was 17 and the oldest 22) and around 20 when they entered post-compulsory education (ranging between 17 and 27). Nearly 60 per cent of individuals in our sample managed to start university or vocational training in the next academic

and it is possible that someone who is unemployed or is a mother will opt for further education in future, this variable is right censored. In proper robustness checks we compare only those who choose university rather than vocational training, restrict the analysis to pupils who graduated from a *Gymnasium*, and then to those who made their educational decision at most one year after *Abitur*.

In contrast to the dependent variable, the independent variables measure characteristics at the age of 17 – i.e. before application is made to university, and usually even before high school is completed. As indicated in the theoretical considerations, we are interested in the impact of perceived probability of success on transition likelihood. So probability of success can be estimated from school performance on the one hand, and from subjective perceptions on the other. Our measure for school performance is obtained by a principal component analysis of school grades in German, mathematics and the first foreign language, which are highly correlated with each other ($grade$).[8] The index is standardized with 0 mean, a standard deviation of 1, and is scaled so that larger values correspond to high performance. Since this is calculated for the whole sample in the Youth Questionnaire, the mean value of 0.21 in our subsample shows that, as would be expected, those who later obtain *Abitur* perform somewhat better at the age of 17 than the others. School grades are computed also as sibling average ($\overline{grade}$) and individual deviation from this average ($\Delta grade$) for subsequent estimations. Table 1 shows the main variables used for the analysis.

As has been mentioned, probability of success might not be estimated exclusively from some easy-to-obtain benchmarks, such as school grades; personal estimations could also be important. We measure expectations of further success through questions about subjectively perceived probability in further education. The following two questions are employed among a set of questions about future events: 'If you think about the future in your career and private life, how probable, in your opinion, is it that the following will occur: i) You receive training or a university place in your preferred field?' ii) 'You successfully finish your training or university studies?'[9] Respondents indicated their answers on a scale from 0 to 100 per cent. Here again, the first principal component is extracted with 0 mean and 1 standard deviation, scaled so that high values correspond to large probabilities of success ($success$). This variable is computed as sibling average ($\overline{success}$) and individual deviation ($\Delta success$), too.

Social class is defined following the Erikson–Goldthorpe–Portocarero scheme (Erikson, Goldthorpe and Portocarero, 1979) and subdivided into two main categories: professionals or managers *vs.* skilled, unskilled and agricultural workers. The reason for using this very simple scheme is that the German educational system is highly selective, and the children of less advantaged social classes are underrepresented among upper-secondary school graduates. Using more detailed classifications, we would not have enough observations in lower-status classes. The definition is based on the father's or the mother's occupation when the individual is 17 (the choice depends on which is better or for which data is available; if both sets of data are missing when the individual is 17 – for example, because of temporary or permanent

year. This rate is in line with the analysis of Federal Statistical Office data by Riphahn and Schieferdecker (2010).

[8] The advantage of using a principal component analysis arises from the fact that, although school grades are highly correlated with each other, they could in some cases display different preferences which might lead to different choices of university subjects. Since we have no information on different university paths, in order to avoid distortions that would result from using the mean of school grades, for example, we combined school grades into a single variable for school performance.

[9] The wording of the questions used to construct the $success$ variable do not differentiate between university and vocational training. To handle this shortcoming, in one of our robustness checks we reduced the sample to those who went to university or embarked on a vocational track. Anyway, we also find a strong correlation between subjectively estimated success and the intention of going to university. These correlations are stronger among pupils with *Abitur*, which could indicate that pupils with *Abitur* more likely had their eyes on a university place when they answered the questions (see Figure A-2 in the Appendix).

unemployment – we looked at the last available occupation). The variable is coded 1 in the case of an advantaged social class (professionals or managers) and 0 otherwise.

Vector Π contains individual-level variables, like gender, year of birth, whether the respondent was on the highest school track at age 17, and survey-year fixed effects. Vector Φ gathers family-level variables. All these variables refer to the point in time when a respondent was 17. Since family characteristics could have changed by the time younger siblings in the same family reach the age of 17, these variables do not necessarily have zero variance among siblings. With net household income, we controlled for the current financial situation, which is considered to be more sensitive to temporary shocks than occupational status. Furthermore, since there are important differences between German federal states in the institutional setting of the school systems, state fixed effects are also employed.

Previous research has shown that spatial distance from a university has an influence on a person's educational decisions (Spieß and Wrohlich, 2008). Indeed, pupils in rural areas often have to commute to university, and other neighbourhood effects in rural settlements could also decrease the probability of choosing to go to university. For this reason, since we cannot control for distance to the nearest university, we include a variable indicating if the type of residency is rural or urban.

It has also been discussed in the past that migration background might be negatively correlated with educational outcomes in Germany. Recent studies have shown, however, that when parental background is controlled for, the effect disappears; this indicates that the most important discrimination experienced by the children of immigrant has to do with disadvantaged socioeconomic status (see, among others, Ludemann and Schwerdt, 2013). We include a control variable for migration background in the regressions, anticipating that it will have no significant impact, since we control for parental occupational background and overall family effects. Controls for the position among siblings and the number of children in the household under age 18 are also included, on the supposition that, all other things being equal, in families with several children, financial resources per capita are lower than in smaller families, and that different parental attention might be devoted to children born later.

Our final sample contains 1,021 individuals from 865 different households for which we have information about grades and subjective perceptions at the age of 17, as well as their choices regarding tertiary education. Furthermore, since the analysis of family effects is crucial for our research question, an extremely interesting feature of the SOEP data is that it allows sibling correlations to be controlled for. Our definition of siblings is based on the identification number of the mother (or of the father, if the former is not available).[10] For the sibling analysis, our sample is reduced to 604 individuals. A specific definition of variables used for the main analysis, the way they are constructed and some descriptive statistics are schematized in Table 1. The principal variables from the SOEP data set are summarized and defined in Table A-1 in the Appendix.

[Table 1 around here]

VI. Descriptive analysis

As we have seen above, past studies have highlighted the fact that the basic reason for low social mobility might lie in access to certain educational tracks. This notion finds further confirmation in a general, descriptive overview of our sample. Figure 1 shows an analysis by *locally weighted regression smoothing* (Cleveland, 1979), which explains the choice of going

[10] Using this algorithm we found the same sibling pairs as provided in a special data set of SOEP which contains information on siblings within the households (BIOSIB). For more information, see SOEP v29 Documentation (SOEP-Group, 2013) pp. 104ff.

to university with school grades and divides the sample by parental background. It is very conspicuous that for most levels of grades (until a relatively high threshold), individuals whose parents are in the upper class are more likely to go to university. Also, the differences between the two classes are particularly remarkable if school performance is relatively low. The difference in the transition ratio gradually decreases and disappears if the grades improve. In other words, if school performance is outstanding (above 2 standard deviation units of mean), status differences should count for less in the transition to university. At average performance, status differences seem to be quite large. It is also noteworthy that in the case of higher social class the association between school grades and the choice of going on to further education is somewhat weaker than in the case of lower social class. This serves as a hint that higher-status pupils decide to go on to further education, rather ignoring their objective probability of success (grades). This highlights the importance of secondary factors in educational decisions.

[Figure 1 around here]

In the next step, we introduced subjectively estimated success into the analysis. The connection of school grades and subjectively estimated success is depicted in Figure 2. Here, the sample is divided by the actual educational choice after *Abitur*. We can see that although there is a certain correlation between the two variables (R^2 is around 0.065, which corresponds to a correlation of around 0.25), individuals in our sample seem to have a very clear view of their future possibilities and/or abilities, since, for each level of grades, those 17-year-olds who later go on to university estimate their future success higher than those pupils who do not. Also, it could be a sign of motivation, self-esteem or optimism. Anyway, it might suggest a positive effect of individually perceived probability of success, driven by a component which is independent of school grades. Moreover, the correlation between school grades and subjectively estimated success is stronger among people who do not go on to university later on. This can be seen especially on the left side of the graph, where pupils who have relatively weak school performance but who later go on university offer a higher estimate of their future success, irrespective of their school grades.

[Figure 2 around here]

Note that, depending on the sample size, the plotted means might not be totally reliable at very large and small grade deviations. Nevertheless, the combination of the evidence obtained so far gives some important clues and paints a picture characterized by light and shadow. On the one hand, if subjective perceptions capture the part of abilities which is not included in grades (objective measure), individual estimations (subjective measure) do matter in determining the transition to tertiary education. On the other hand, for a given (subjective and objective) skill level, social class somehow influences people's chances of educational success. The regression analysis in the following sections will verify the suggestions given by this first descriptive evaluation and will deepen the insights into the driving forces of the underlying mechanism.

VII. Results of the econometric analysis

Table 2 displays the results of the multivariate analysis described in Section IV, where standard errors are robust, clustered among siblings. As we can see in the results of simple OLS and Logit (Columns 1 and 2), both grades and subjectively estimated success are associated positively with the choice to go to university. The average marginal effects of the binary Logit analysis (Column 2) show that one standard deviation increase in grades improves the likelihood of choosing to go on to university by some 9 percentage points, while the figure is 4 percentage points in the case of subjectively estimated success. The results from these two first columns offer some suggestion that the probability of success is not estimated exclusively from school performance, but that subjective estimations play a crucial role in the educational transition process. These first results could, however, be biased by unobserved family heterogeneity.

For this reason, as explained above, random effects Logit is performed on the full sample. The results of this analysis (Column 3) show that parameter estimations are naturally very close to the Logit model. This is confirmed in the case of subjective perceptions by the results using Mundlak's approach, reported in Column 4. First, subjectively estimated success remains significant; second, the Mundlak test, which measures the joint significance of group means, hints at an unbiased random effects estimator (confirmed by a Hausman test of Model 3). Interestingly, school grades are no longer significant – a finding that will be discussed below. Finally, Columns 5 and 6 show the results of the estimations where we restricted our sample to those who have siblings in the data. First, we considered all individuals who have at least one sibling in the data (Model 5), regardless of whether all siblings in one family achieved *Abitur* and decided to go on to further education afterwards. Then, in the second restriction, we considered only those families where all siblings achieved *Abitur* (Model 6).

Considering school grades, we see that only the sibling average appears to be a significant predictor of the choice to go on to university (not the individual deviation from it, which seems to have no independent effect). This result could show that the well-known feature of the German educational system in the early stages of schooling – namely that family background plays a crucial role in school performance – is replicated at the transition to tertiary education. Another interesting pointer from these results is that in families where the performance of children is relatively strong, individual deviations are not important: in families with at least one child at university, less talented siblings are anyway *pushed* to choose the academic path. The same is true of families where overall school performance is relatively low: in such cases, individual talent seems not to be supported when it comes to the choice of going on to tertiary education. Fitting this result to the framework proposed by Gambetta (1987), we could argue that siblings with outstanding individual performance are not able to *jump.* This claim is underlined by the observed sibling correlation of 0.275 in Model 6.[11]

However, there may be some hope lurking behind our results: it would seem that *push–jump mechanisms* work differently in the case of subjectively estimated success. While families with a high average rate of estimated success among siblings have a higher probability of a child entering university, individual deviations are still significant for the educational choice. An individual who is more optimistic than his/her siblings when it comes to estimating personal future success has a higher likelihood of going to university. Thus, this could be a possible jumping point with special importance. Since it is very difficult to influence family-level characteristics, policy interventions might be more effective if they are targeted at individuals. However, the great similarities between siblings in the forces influencing educational attainment – shown here also by the sibling correlation, which is approximately one-quarter of total variance – highlights the importance of family in this educational decision.[12]

[11] The sibling correlations in permanent earnings in Germany is found by Schnitzlein (2013) to be 0.432 for brothers and 0.391 for sisters.

[12] Note that interpretations of family averages in school grades and subjectively estimated success are only valid if these averages are not correlated with unobserved family factors. Otherwise, the coefficients of the cluster averages would be inconsistent. Controlling for parental social status might mitigate this risk to some extent.

Parental occupational class position does have an independent effect on educational decisions, even at later stages in the educational career. Its impact might also partly be captured by school grades and the school type dummy. It is very likely that parental background is more important in earlier educational decisions, as Boudon (1974: 30) points out, because pupils are more influenced by their parents' social background at earlier stages than later on. In this sense, our results might indicate that earlier educational decisions could have long-lasting consequences, since going to university is only an option if someone has already managed to obtain the *Abitur*. Compared to class position, financial constraints – measured by household income – are not a significant predictor of this late educational choice. This could be a consequence of relatively generous funding and grant opportunities at German universities.[13] Finally, age at the time of taking the high-school final exam also contributes to the transition rate: apparently those who achieve *Abitur* later in life are more motivated to go to university.

[Table 2 around here]

Summarizing the results from the regression analysis, we find that school grades are crucial in deciding to go on to university after taking the high-school final exam. This is not very surprising, since it reflects the structure of the German educational system, where good grades are admission tickets to tertiary education. What is remarkable is that individual deviation from the sibling average has no independent effect. Thus, pupils seem to be *pushed* by family factors when opting for further education. This might confirm the widely recognized fact that parental background plays a crucial role in mitigating equality of educational opportunities in Germany (e.g. Dustmann, 2004).

Another key finding is that subjective perceptions seem to have an individual effect even after controlling for family background factors. In our view, our analysis controls for the overall effect of family background better than other analyses, which use only parental occupational, educational or cultural status as a proxy. Overall family effects (including nature and nurture) are kept constant with sibling averages (and with shared variance among siblings in random effect models), and school performance is controlled for by school grades and type of secondary school. Since the established (positive) impact of subjectively estimated success seems to be independent of social status (as well as of school performance and overall family effects), it could be interpreted as a point of *jumping off* from the selective and structured German education system. This gives space to interpretations from two different perspectives: first, subjectively estimated success might capture some non-measured cognitive ability that is not reflected in school grades and does not originate in the family; or second, it might be the impact of some non-cognitive skill, like optimism or positive thinking.

At this stage, our results could be shaped by unobserved school factors or other sources of endogeneity. Even though we controlled for the type of school, data restrictions mean that we cannot control for unobserved school-level heterogeneity, peer-group effects or school quality. We believe that school factors should be captured by family factors, which we control for. But to be sure, in the following sections we discuss this issue further and present our robustness checks.

Still, the interpretations of individual deviations are unaffected, since these are uncorrelated by construction with the error term.

[13] The monthly amount of the so-called BAföG system was between 350 and 450 Euro in the observed time period (Federal Statistical Office).

VIII. Dealing with omitted variables

As Figure 2 indicates, there is a remarkable correlation between school grades and subjectively estimated success. Thus, multicollinearity could lead to biased standard errors, especially if the sample size is small. Furthermore, in our particular case a possible source of endogeneity could be the fact that the correlation between $grade$ and $success$ might be driven by unobserved school quality: in good schools, pupils might have additional information about their talent – namely that they are attending a good school – and consequently may predict their future success higher at each level of school grades (even if this level is more difficult to attain than in other schools). We cannot really examine this hypothesis, since we lack information on school quality. However, we observe that the strength of correlation varies according to the school type attended at age 17. As shown in Figure A-1 in the Appendix, in more competitive school types (*Gymnasium*) the correlation is higher than in others which have no direct path leading to university (*Realschule*). This evidence could confirm our suppositions about omitted variable bias.

We also applied an alternative approach to dealing with this problem. Instead of school grades, we used an alternative measure of ability, based on a test of verbal, numerical and figural intelligence.[14] This measure is, by definition, not an outcome of the educational system, and the correlation of subjectively estimated probability of success with ability is weaker than in the case of school grades (0.18 instead of 0.25, while the correlation between school grades and ability is 0.34). Controlling for ability, the impact of subjectively estimated success remains significant (see Table 3). So we can conclude that even if we could somehow fix unobserved school quality, subjectively estimated success is a meaningful predictor of the educational choice. Unfortunately, our measure of ability is only available after 2006. Hence, sample size is very limited and sibling analysis is impossible, therefore this specification is only applied on *Basic Models.*

[Table 3 around here]

A further possible way of dealing with omitted variable bias would be to find an instrument that correlates to our probability of success variable, but not directly to the choice to go on to tertiary education. We identified 'positive thinking about life' as such an instrument (which depends on issues like self-estimated probability of finding employment, of marrying, etc.; see definition in Table A-2 in Appendix). It should be remembered that probability of success measures a specifically *educational* positive outlook. Although there should be no doubt that education-specific and general positive thinking correlate with each other, in educational decisions (like opting for a university education) education-specific positive outlooks should be more important. On the other hand, overall positive thinking has the advantage of being less sensitive to influences from school quality. If someone is optimistic by nature, it should matter less what kind of feedback he or she receives from school in the form of school grades. As Table 4 indicates, instrumenting perceived probability of success, its impact remains significant whether 2SLS (Model 2) or IV-Probit (Model 3) is applied. Its size is somewhat greater than the parameters from the base models in Table 2.

[Table 4 around here]

A final issue that needs to be clarified has to do with the robustness of our measurement of subjective perceptions of future success. As a first step, and leaving space for further

[14] The measure used is an I-S-T 2000R test contained in the SOEP data (Amthauer, Brocke, Liepmann and Beauducel, 2001). The first principal component is extracted from the numbers of correct answers of the three subscales. Details about the variables: Table A-1 in Appendix.

research, we try to identify the relationship between subjective perceptions about future success and some well-known personality traits. As discussed above, previous sociological and economic literature has shown increasing interest in psychological measures. Special attention is focused on *locus of control* (Rotter, 1966) – the belief in one's ability to control life – and the well-known Big Five personality traits (Costa and McCrae, 1995) of openness, conscientiousness, extraversion, agreeableness and neuroticism. Figure 3 shows the two-way connection between these personality measures (applied in SOEP: Weinhardt and Schupp, 2011) and our measurement of subjectively estimated success for the whole youth sample. We see that subjectively estimated success has a positive connection with positive personality traits like openness, conscientiousness, extraversion, agreeableness and internal locus of control; and a negative relationship with negative traits like neuroticism and external control. The relationship is remarkable, having the highest correlation with conscientiousness (0.298) and the lowest with neuroticism (-0.124).

[Figure 3 around here]

Based on the correlation between these variables, one concern could be that the relationship between subjectively estimated success and the likelihood of going on to university is spurious, and what actually explains the transition to tertiary education are some of the well-known personality measures. Because of multicollinearity, to include both subjectively estimated success and personality measures would not be a possible solution (especially due to the sample size: as mentioned before, personality variables have only been available since 2006 in SOEP). An alternative way of testing this hypothesis is to replace subjective estimated success with the personality measurements cited above, perform the estimations and evaluate the coefficients. For the same reason as before, we also cannot perform sibling analysis and so restrict this technique to *Basic* and *Family Models*. Figure 4 shows the estimated coefficients with a 95 per cent confidence interval, where subjectively estimated success is replaced by the Big Five personality measures (Graph A) and locus of control (Graph B). All the other variables are the same as in the main analysis (see Table 2). As the graph shows, the confidence intervals of the estimated parameters cross the zero line in the case of the Big Five. In the case of locus of control, external control (life events are influenced by fate) is not significant, while internal control (life events can be influenced personally) seems to be a significant predictor for the choice of tertiary education. We conclude that we cannot entirely exclude the possibility that subjectively estimated success is only important in the transition to tertiary education because it captures the impact of some personality traits; but we have established that personality variables perform less effectively in explaining the choice of going on to tertiary education than does our measurement of subjectively estimated success. This suggests that subjectively estimated success captures more information than the observed personality traits.

[Figure 4 around here]

IX. Handling selectivity and other robustness checks

In their seminal analysis, Cameron and Heckman (1998) argue that since educational decisions are consequences of previous educational decisions, observed and unobserved components influencing former educational decisions have an influence on subsequent decisions as well. In our particular case, unobservables might bias the parameter estimations if those individuals who do not achieve *Abitur* have significantly different characteristics from those who do. For example, pupils with low school performance (probably from lower-status families) should be underrepresented in our sample, as they are underrepresented in upper-secondary school types in Germany. Therefore, those who come from a disadvantaged background and who – against all the odds – achieve *Abitur* and then apply to tertiary education might have an increasingly positive set of unobserved characteristics (Riphahn and Schieferdecker, 2010).

We accounted for this, applying a Heckman two-stage procedure (Heckman, 1979) using a set of variables in the first stage, which should influence the possible achievement of *Abitur*, but not the choice to go on to tertiary education. These variables recognize whether the school type recommended by the teacher after primary education was a *Gymnasium* ($recc\_gymn$), the fact of a pupil being at a *Gymnasium* at the age of 17 despite having a lower teacher recommendation ($gymn\_norec$) and if a sibling has yet achieved *Abitur* (sib_abi). The results of the first stage (Table A-2 in the Appendix) show that those who were recommended to choose a *Gymnasium* after elementary school, those who were at a *Gymnasium* in spite of the negative school recommendation, and those whose sibling had achieved *Abitur* are all more likely to be included in the sample (i.e. they also achieved *Abitur*). Also remarkable is that the impact of subjectively estimated future success is negative at the first stage. One possible explanation for this could be that the questions used primarily to construct this variable do not differentiate between vocational training and university places. And so we are not able to determine which kind of further education was considered by the respondent. This should be less problematic when analysing educational decisions after *Abitur*, but at that level everybody is still considered – even those who are in the lowest secondary school track and are only able to opt for vocational training. Another possible explanation for the negative sign of the parameter could be the omitted school quality measure already discussed. Note that this sample includes those pupils who are in lower secondary school tracks and have never managed to switch onto a higher track; and also that the items used to construct the success variable were required from every respondent at the age of 17, regardless of his or her objective chances of going on to further education. For instance, in less competitive schools it might be easier to attain better grades. So, while school quality increases the probability of a successful transition to tertiary education, it decreases self-esteem because of relatively poor grades (the mechanism is explained by Marsh and Hau, 2003). At the second stage of the Heckman procedure – after correcting for unobservables in the selection mechanism – the main conclusions do not change. The lack of significance in the inverse Mills ratio can be interpreted as no bias in the estimated parameters at the second stage, even though there is evidence of selection from the first stage of the model (see Table A-2 and Table A-3 in the Appendix).

There is another possible way of sample selection. Since university entry can only be established for those who have remained in the panel, we might imagine that those who have changed their place of residence on admittance to university in a different city are more likely to be missing from the sample. In reality these adolescents entered university, but since they did not stay in the panel, we cannot 'establish' their university admittance. If these missing pupils are the best motivated, with good school performance, our estimations might be biased. In fact, however, stability (defined as the difference between actual and previous household ID) has a positive effect on whether people drop out of the panel; moreover, school grades and subjectively estimated success do not play a significant role in the multivariate analysis explaining panel dropout, and so this argument is not supported by the data.

As further robustness checks, we changed some definitions and restricted the sample. A comparison between the estimated parameters for $success$ and $\Delta success$ from the original model and the robustness checks are plotted in Figure 5. Graph A summarizes the estimated parameters from Table 2, which is the original model. On Graph B, the sample is restricted to a more homogeneous subsample – those who graduated from a *Gymnasium*. Note that in Germany, attendance at a university of applied science (*Fachhochschule*) is also possible with a *Fachabitur*. Moreover, an *Abitur* can also be achieved from less competitive integrated schools (*Gesamtschule*).

In Graphs C, D and E, the problem to be addressed is the right-censored nature of the data. Failure (not being admitted to the university) could only be established for those who 'survived' in our panel survey.[15] In the case of those who dropped out of the panel, we do not actually know whether they managed to enter university later. To deal with this problem, one possible solution is to restrict the sample to those who made some decision after *Abitur*, since in the case of this group we know that respondents survived in the panel until they decided one way or the other. In Graph C, the dependent variable is 1 if someone went to university and 0 in the case of vocational training after *Abitur*, so those pupils who made any kind of educational decision after the high-school final exam are analysed. In Graph D, we considered only those who made the educational choice in the same year as *Abitur* – or at most one year after it. Note that by definition the sample contains (in this case as well) those who chose between university and vocational education after the high-school final exam. This kind of restriction seems to be necessary, since the majority of the transitions to university occur straight after *Abitur* (Statistisches Bundesamt, 2014: 145), and therefore the 'success likelihood' is not equally distributed.[16] Another possible way of dealing with panel erosion is by restricting our sample to those who 'survived' in the panel for at least two years after *Abitur* (Graph E). Contrary to the previous two settings – where the choice between university and vocational training is analysed – here the choice of going on to university is contrasted with any other options in a two-year timeframe after the high-school final exam. If somebody is coded as 0, we know that that person did not manage the transition; and because everybody 'survived' in the panel, the bias caused by wrong classification because of panel attrition is minimized.

Lastly – in Graph G – we changed the $status$ variable from social background based on occupation to social background based on education: 1 if at least one of the parents has a university degree and 0 otherwise. The reason behind this last change is that parents with a university degree might be more likely to push their children to obtain a degree if, for example, holding a degree is a kind of family 'tradition'. Applying all these modifications, we obtain very similar results in terms of direction, sign and significance of the parameter of interest, underlining the robustness of the analysis performed.

X. Conclusions

In this study we have addressed the question of whether the effect of parental background is predominant in educational choices, or if individual contributions might still play a significant role. In particular, we evaluated the effect of subjective perceptions of future success on the transition to tertiary education, controlling for the overall influence of family background through a sibling analysis. The theoretical basis for our approach derives from rational choice models applied in educational sociology, which regard probability of success as one of the major components in the explanation of educational transitions.

Our initial point of this analysis was that perceived probability of success might not be obtained exclusively from school grades: unobserved factors might also shape this perception. And so we used subjective perceptions about future educational success to measure this

[15] However right-censoring derives also from panel erosion. This issue is not the same as sample selectivity, explained above. Whereas in the case of sample selectivity the question is whether those who remained and those who dropped out from the panel have the same characteristic, in the case of right-censoring the problem is that we have a specific observation window, and respondents could opt for university education outside this time period as well.

[16] The graphical results presented in this graph are reinforced by survival analysis (results are available from the authors on request).

usually unobserved component. Moreover, we assumed this subjective estimation to be influenced by family 'endowments' – taken not just to be parental income, occupational or educational class position, but also norms, values or parental wishes transmitted within the family – and ruled out the overall influence of this unobserved family heterogeneity by sibling analysis. Hence, we were able to control for factors shared by siblings – and that presumably originate within the family.

Our findings show that among those who obtain the highest secondary-school diploma (and controlling for parental background), those 17-year-olds who put a higher estimate on their future success are later more likely to choose university education. This result is robust even after controlling for school performance, school type, abilities and – as far as the data allows – school type and personality traits. Nevertheless, our findings show another aspect, too: grades, which are intended as an objective measure of school attainment, certainly have a crucial impact on a pupil's further educational career, but – it would seem – not at the individual level. Children from families with higher grade averages among siblings indeed do have a higher probability of going to university, regardless of their individual grade level. The same mechanism works the other way around: outstanding children from disadvantaged families (i.e. with lower average school grades among siblings) have no significant positive chance of going to university. This result highlights the fact that in Germany the strong selection by family background at secondary school (as found by other studies) seems to recur at the transition to tertiary education.

Still, we confirmed that – following the framework of rational choice models – perceived probability of success is an important driver in educational decisions. Our findings provide ample evidence for considering that some individual characteristics – namely those that capture subjective perceptions of future success – might have an independent effect and work as a *jumping point* from the selection mechanism of the educational system. Furthermore, we established two interesting features that should play an important role. First, probability of success should not be exclusively driven by easily obtained signs (such as school grades): subjective perceptions also play an important role. Second, in contrast to school grades, such subjective estimations seem to be independent of family background. In terms of policy recommendations, one implication of this could be that the social planner should foster methods that contribute to better self-knowledge among young individuals, so that they can discover latent skills that could be important in the choice of knowledge-intensive educational scenarios. However, before precise proposals can be formulated, the exact nature of subjective perceptions of future success has to be investigated further.

One possible interpretation resulting from our findings is that, being independent of school performance and family background, subjective perceptions might mirror some non-cognitive ability. Another possibility is that subjective perceptions might display some unobservable cognitive ability, like perspicacity, allowing future events to be assessed very well. We have furthermore tried to illustrate whether and how strongly our measurement of subjective perceptions is associated with abilities and certain personality traits, and have shown that there are no clear patterns with the available measurements. More profound analysis of the subject should be undertaken by further research, for example into the relationship between subjective perceptions and the passion for long-term goals, namely *Grit* (see Duckworth et al., 2007), which is currently attracting animated discussion in the psychological literature.

For the study of the role of the individual component in educational decisions, our findings provide interesting clues. We have established that, apart from objective indicators of success (like school grades), subjective perceptions are also important for educational transitions and that these seem to be independent of both background effects and abilities. The question of whether these perceptions are driven by some non-measured cognitive or non-cognitive skills leaves space for further research. Moreover, we found that our results are stable if we control for cognitive skills, and that personality traits themselves do not have the same explanatory power for educational decisions as subjectively estimated success. These promising results could thus imply that optimism, awareness of one's own talents and self-esteem have an important role in educational choices. Future research should, however, clarify how these subjective perceptions might be developed, and how an increase in subjectively estimated success could translate into a higher probability of transition to university, thus improving the equality of educational opportunities. At this stage we can only indicate that this stream of research could have interesting policy implications, such as encouraging pupils to recognize their strength and help them to build confidence based on their own talents. Finding individual-level characteristics that are independent of family factors has the advantage that encouragement at the individual level is much easier than prescribing regulations for families.

References

Akerlof, G. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, *65*(5), 1005–1027.

Amthauer, R., Brocke, B., Liepmann, D. and Beauducel, A. (2001). *I-S-T 2000R. Intelligenz-Struktur-Test 2000R*. Göttingen: Hogrefe.

Anger, S. (2012). Intergenerational transmission of cognitive and noncognitive skills. In J. Ermisch, M. Jäntti and T. Smeeding (eds), *From Parents to Children: The intergenerational transmission of advantage*. New York: Russell Sage Foundation.

Baker, D. and Stevenson, D. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. *Sociology of Education*, *59*, 156–166.

Bauer, P. and Riphahn, R. (2006). Timing of school tracking as a determinant of intergenerational transmission of education. *Economics Letters*, *91*(1), 90–97.

Bauer, P. and Riphahn, R. (2009). Age at school entry and intergenerational educational mobility. *Economics Letters*, 3977.

Becker, G.S.G. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy*, *87*, 1153–1189.

Björklund, A. and Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, *19*(4), 465–474.

Boudon, R. (1974). *Education, Opportunity and Social Inequality*. New York: Wiley-Interscience.

Bratti, M., Checchi, D. and Blasio, G. De. (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. *Labour*, *22*(s1), 53–88.

Breen, R. and Goldthorpe, J.H. (1997). Explaining educational differentials: Towards a formal rational action theory. *Rationality and Society*, *9*(3), 275–305.

Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*(52), 781–861. doi:10.1111/j.1468-0327.2007.00189.x

Brunori, P., Peragine, V. and Serlenga, L. (2012). Fairness in education: The Italian university before and after the reform. *Economics of Education Review*, *31*(5), 764–777. doi:10.1016/j.econedurev.2012.05.007

Cameron, S.V. and Heckman, J.J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy*, *106*(2), 262. doi:10.1086/250010

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836.

Corneo, G. and Jeanne, O. (2010). Symbolic values, occupational choice, and economic development. *European Economic Review*, *54*(2), 237–251. doi:10.1016/j.euroecorev.2009.04.006

Costa, P.T. and McCrae, R.R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, *64*(1), 21–50. doi:10.1207/s15327752jpa6401_2

Davies, R., Heinesen, E. and Holm, A. (2002). The relative risk aversion hypothesis of educational choice. *Journal of Population Economics*, *15*(4), 683–713. doi:10.1007/s001480100087

Duckworth, A.L., Peterson, C., Matthews, M.D. and Kelly, D.R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–101. doi:10.1037/0022-3514.92.6.1087

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, *56*(2), 209–230.

Erikson, R. and Jonsson, J.O. (1996). *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, CO: Westview Press.

Erikson, R., Goldthorpe, J.H. and Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology*, *30*(4), 415–441.

Esser, H. (1999). *Soziologie. Spezielle Grundlagen.* Band 1: *Situationslogik und Handeln*. Frankfurt/New York: Campus Verlag.

Falch, T. and Strøm, B. (2013). Schools, ability, and the socioeconomic gradient in education choices. *Journal of Behavioral and Experimental Economics* (formerly *Journal of Socio-Economics*), *43*, 49–59. doi:10.1016/j.socec.2013.01.003

Gambetta, D. (1987). *Were They Pushed or Did They Jump?: Individual decision mechanisms in education*. Cambridge: Cambridge University Press.

Hanushek, E.A. and Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal*, *116*(510), C63–C76.

Hartlaub, V. and Schneider, T. (2012). Educational choice and risk aversion: How important is structural vs. individual risk aversion? *SOEP Papers on Multidisciplinary Panel Data Research*, *433*, 1–41.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161. doi:10.2307/1912352

Heckman, J.J., Stixrud, J. and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*.

Heineck, G. and Riphahn, R.T. (2009). Intergenerational transmission of educational attainment in Germany – the last five decades. *Journal of Economics and Statistics*, *229*(1), 36–60.

Keller, S. and Zavalloni, M. (1964). Ambition and social class: a respecification. *Social Forces*, *43*(1), 58–70.

Lohmann, H. and Witzke, S. (2011). BIOEDU (beta-Version): Biographical data on educational participation and transitions in the German Socio-Economic Panel Study (SOEP). *DIW Data Documentation*, *58*.

Ludemann, E. and Schwerdt, G. (2013). Migration background and educational tracking. *Journal of Population Economics*, *26*(2), 455–481. doi:10.1007/s00148-012-0414-z

Marsh, H.W. and Hau, K.-T. (2003). Big-Fish–Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. doi:10.1037/0003-066X.58.5.364

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85. doi:10.2307/1913646

Murphy, J. (1981). Class inequality in education: Two justifications, one evaluation but no hard evidence. *British Journal of Sociology*, *32*(2), 182–201. doi:10.2307/589445

Murphy, J. (1990). A most respectable prejudice: inequality in educational research and policy. *British Journal of Sociology*, *41*(1), 29–54. doi:10.2307/591017

Neugebauer, M. and Schindler, S. (2012). Early transitions and tertiary enrolment: The cumulative impact of primary and secondary effects on entering university in Germany. *Acta Sociologica*, *55*(1), 19–36. doi:10.1177/0001699311427747

OECD. (2010). *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes* (Volume II). OECD Publishing. doi:10.1787/9789264091504-en

Oehlert, G.W. (1992). A note on the Delta Method. *American Statistician*, *46*(1), 27–29. doi:10.2307/2684406

Paulus, W. and Blossfeld, H.-P. (2007). Schichtspezifische Präferenzen odersozioökonomisches Entscheidungskalkül? *Zeitschrift für Pädagogik 53*, *53*(4), 491–508.

Pietsch, M. and Stubbe, T. (2007). Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal*, *6*(4), 424. doi:10.2304/eerj.2007.6.4.424

Raudenbush, S. and Bryk, A. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*(1), 1–17.

Raudenbush, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and data analysis methods*. Sage.

Riphahn, R.T. and Schieferdecker, F. (2010). The transition to tertiary education and parental background over time. *Journal of Population Economics*, *25*(2), 635–675. doi:10.1007/s00148-010-0347-3

Roemer, J.E. (2000). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Rosen, B.C. (1956). The Achievement Syndrome: A psychocultural dimension of social stratification. *American Sociological Review*, *21*(2), 203–211.

Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcements. *Psychological Monographs*, *80*(1), 1–28.

Sacerdote, B. (2011). Nature and nurture effects on children's outcomes: What have we learned from studies of twins and adoptees? In *Handbook of Social Economics* (Vol. 1). Elsevier B.V. doi:10.1016/B978-0-444-53187-2.00001-2

Schindler, S. and Reimer, D. (2010). Differentiation and social selectivity in German higher education. *Higher Education*, *61*(3), 261–275. doi:10.1007/s10734-010-9376-9

Schneider, T. (2008). Social inequality in educational participation in the German school system in a longitudinal perspective: Pathways into and out of the most prestigious school track. *European Sociological Review*, *24*(4), 511–526. doi:10.1093/esr/jcn017

Schnitzlein, D.D. (2013). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, *27*(1), 69–89. doi:10.1007/s00148-013-0468-6

Schütz, G., Ursprung, H. and Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, *61*(2), 279–308.

SOEP-Group (2013). *SOEP 2012-Documentation on biography and life history data for SOEP v29*.

Solon, G., Corcoran, M., Gordon, R. and Laren, D. (1991). A longitudinal analysis of sibling correlations in economic status. *Journal of Human Resources*, *26*(3), 509–534.

Spieß, C.K. and Wrohlich, K. (2008). Does distance determine who attends a university in Germany? *SOEP Papers on Multidisciplinary Panel Data Research*, 118.

Statistisches Bundesamt (2014). *Statistisches Bundesamt Bildung und Kultur Nichtmonetäre hochschulstatistische Kennzahlen* (Vol. 49, pp. 1–681). Wiesbaden: Statistisches Bundesamt.

Stocké, V. (2007). Explaining educational decision and effects of families' social class position: an empirical test of the Breen Goldthorpe Model of educational attainment. *European Sociological Review*, *23*(4), 505–519. doi:10.1093/esr/jcm014

Stocké, V. (2010). Der Beitrag der Theorie rationaler Entscheidung zur Erklärung von Bildungsungleichheit. In G. Quenzel and K. Hurrelmann (eds), *Bildungsverlierer. Neue Ungleichheiten*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Tamm, M. (2008). Does money buy higher schooling?: Evidence from secondary school track choice in Germany. *Economics of Education Review*, (55), 1–26.

Tolsma, J., Need, A. and Jong, U. De (2010). Explaining participation differentials in Dutch higher education: The impact of subjective success probabilities on level choice and field choice. *European Sociological Review*, *26*(2), 235–252. doi:10.1093/esr/jcp061

Uhlig, J., Solga, H. and Schupp, J. (2009). Inequality in educational opportunities: underachievement and the role of personality traits. *Zeitschrift für Soziologie*, *38*, 418–440.

Wagner, G.G., Frick, J.R. and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – scope, evolution and enhancements. *Schmollers Jahrbuch : Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften*, *127*(1), 139–169.

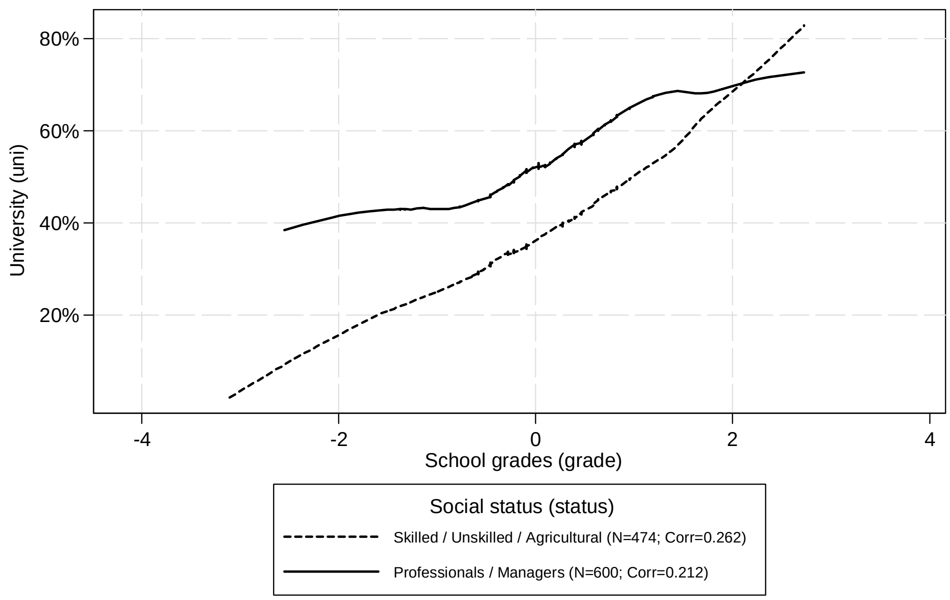
Weinhardt, M. and Schupp, J. (2011). Multi-Itemskalen im SOEP Jugendfragebogen. *DIW Data Documentation*, *60*.

Woessmann, L. (2008). How equal are educational opportunities? Family background and student achievement in Europe and the United States. *Zeitschrift für Betriebswirtschaft*, *78(1)*(1284), 45–70.

Wooldridge, J.M. (2009). *Introductory Econometrics: A modern approach* (4th edition). Mason, OH: South-Western.

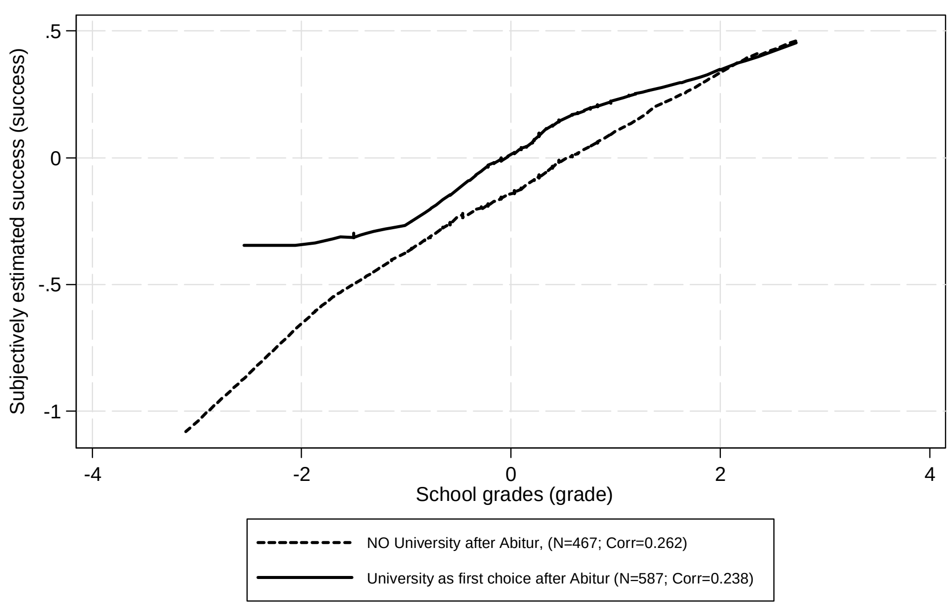
Figures and Tables

Figure 1: The connection between school grades and the choice to go to university, by parental background – locally weighted regression smoothing



Population: those who took high school final exam (Abitur)
Method: Lowess Smoothing
Data: SOEP, v29, N=1074

Figure 2: The connection between school grades and subjectively estimated success, by later educational decision – locally weighted regression smoothing



Population: those who took high school final exam (Abitur)
Method: Lowess Smoothing
Data: SOEP, v29, N=1054

Figure 3: The connection between various kinds of personality variables and subjectively estimated success

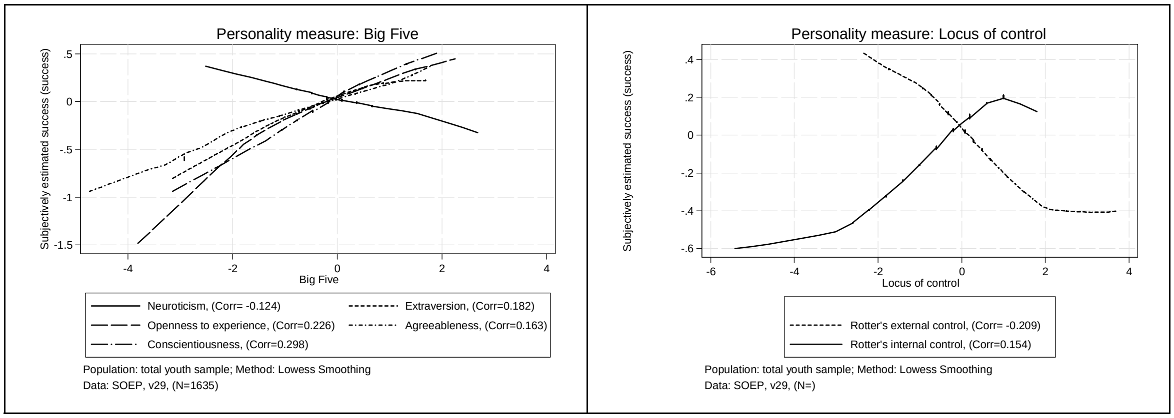


Figure 4: Robustness checks, using personality measures instead of subjectively estimated success

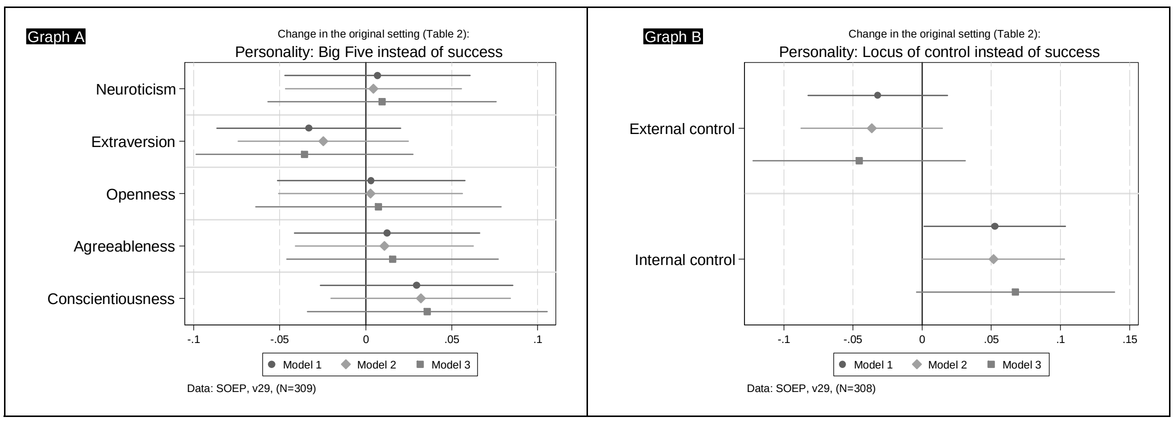


Figure 5: Robustness checks, the estimated parameters of success and Δsuccess in different settings

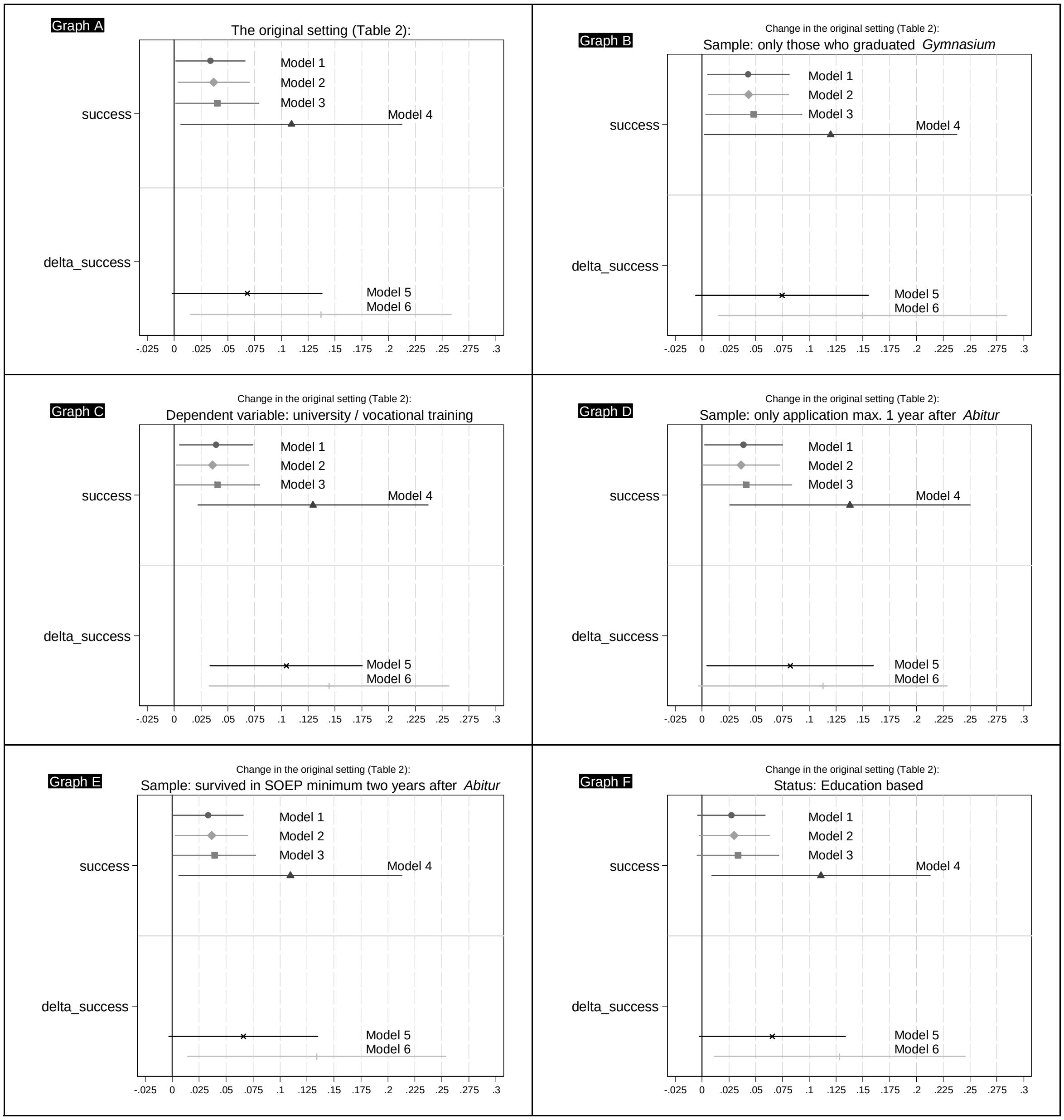


Table 1: Definition of variables used in the main analysis, N=1021

	Definition	Construction	Range min./max.	Mean/(sd)
Dependent variables				
uni	Chose university (Sample: those who made *Abitur*)	The first choice is calculated	0/1	0.48/(0.50)
Independent variables				
Probability of success, personal estimations				
$success$	Subjectively estimated success Questions used: If you think about the future in your career and private life, how probable, in your opinion, is it that the following will occur: - *You receive training or a university place in your preferred field?* - *You successfully finish your training or university studies?*	Principal component analysis	-3.92/1.38	-0.02/(0.86)
$\overline{success}$	Subjectively estimated success; sibling average	$\frac{\sum_{sibling=1}^{sibling=n} success}{N_{sibling}}$	-3.92/1.38	0.00/(0.74)
$\Delta success$	Subjectively estimated success; difference from sibling average	$success - \overline{success}$	-2.15/2.27	-0.02/(0.46)
Probability of success, derived from school performance				
$grade$	School grades in German, mathematics and first foreign language	Principal component analysis	-2.55/2.73	0.21/(1.00)
$\overline{grade}$	School grades; sibling average	$\frac{\sum_{sibling=1}^{sibling=n} grade}{N_{sibling}}$	-2.55/2.73	0.21/(0.86)
$\Delta grade$	School grades; difference from sibling average	$grade - \overline{grade}$	-2.12/2.12	0.01/(0.52)
Parent's social status				
$status$	Social status Erikson, Goldthorpe and Portocarero, (1979) class scheme - *professionals or managers (1)* - *skilled, unskilled and agricultural workers (0)*	The best score (mother/father) is used. If missing, the score from last wave is used.	0/1	0.56/(0.50)
Π Vector of individual-specific variables				
$male$	Respondent is male (ref. female)		0/1	0.45/(0.50)
$birth$	Year of birth		1982/1992	1987/(2.70)
$gymn$	Respondent is in Gymnasium at age 17 (ref. any other school type)		0/1	0.74/(0.44)
abi_age	Age at *Abitur*		17/23	19.02/(1.07)
$year$	Year of the survey	Dummy variables	2000/2009	
Ψ Vector of family-specific variables				
ι_net_hinc	Net household income	Inflated to 2010 price level[17]	7.62/12.21	10.77/(0.48)
$state$	Federal state of residence	Dummy variables	1/16	
$rural$	Residence in rural region (ref. urban or urbanized region)		0/1	0.25/(0.43)
mig	At least one of the parents has migration background		0/1	0.17/(0.38)
pos_sib	Birth order	Position among siblings according to year of birth	1/8	1.62/(0.802)
$children$	Household members under 18		0/8	1.74/(0.94)

[17] http://www.zinsen-berechnen.de/inflation/tabelle-inflationsrate.php

Table 2: Regression results, dependent variable uni, average marginal effects

		Basic models		Family models		Sibling models	
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		OLS	Binary Logit	Random effect Logit	Mundlak's Random eff.	Sibling random effect Logit	Sibling+Abi rand. eff. Logit
	$success$	0.034**	0.037**	0.040**	0.109**		
		(0.016)	(0.017)	(0.020)	(0.053)		
	$\Delta success$					0.068*	0.136**
						(0.036)	(0.062)
	$\overline{success}$					0.076**	0.123**
						(0.035)	(0.057)
	$grade$	0.089***	0.089***	0.096***	0.042		
		(0.016)	(0.015)	(0.017)	(0.038)		
	$\Delta grade$					0.031	0.042
						(0.030)	(0.042)
	$\overline{grade}$					0.107***	0.115***
						(0.031)	(0.043)
	$status$	0.083**	0.083***	0.089**	-0.064	0.120**	0.068
		(0.033)	(0.031)	(0.035)	(0.151)	(0.048)	(0.079)
Π	$male$	0.035	0.038	0.043	0.033	0.021	0.027
		(0.029)	(0.029)	(0.032)	(0.071)	(0.041)	(0.065)
	$birth$	0.003	0.004	0.012	0.076	0.086	-0.004
		(0.044)	(0.041)	(0.044)	(0.093)	(0.063)	(0.016)
	$gymn$	0.195***	0.209***	0.228***	0.336***	0.279***	0.360***
		(0.035)	(0.034)	(0.040)	(0.113)	(0.050)	(0.085)
	abi_age	0.092***	0.096***	0.104***	0.051	0.077***	0.104***
		(0.014)	(0.014)	(0.016)	(0.042)	(0.022)	(0.035)
	$year$	Yes	Yes	Yes	Yes	Yes	Yes
Ψ	ln_net_hinc	0.034	0.035	0.039	0.070	0.018	0.007
		(0.036)	(0.035)	(0.038)	(0.171)	(0.053)	(0.087)
	$rural$	-0.058	-0.054	-0.059	-0.455	-0.127*	-0.028
		(0.050)	(0.047)	(0.052)	(0.472)	(0.075)	(0.084)
	mig	0.049	0.051	0.058	0.054	0.089	0.024
		(0.041)	(0.042)	(0.046)	(0.042)	(0.061)	(0.110)
	pos_sib	0.021	0.019	0.021	0.144	-0.010	0.124**
		(0.020)	(0.019)	(0.021)	(0.091)	(0.027)	(0.056)
	$children$	0.001	0.002	0.002	0.050	0.006	0.039
		(0.018)	(0.017)	(0.017)	(0.112)	(0.023)	(0.059)
	$state$	Yes	Yes	Yes	Yes	Yes	Yes#
Mean of indep. var.					Yes		
Constant		Yes	Yes	Yes	Yes	Yes	Yes
Observations		1,021	1,021	1,021	1,021	604	232
No. of families		865	865	865	865	448	111
R-squared		0.223					
F-stat		13.25***					
chi2			170.5***	70.67***	295.5***	50.10*	27.38**
Log likelihood			-574.3	-573.2		-334.6	-125.3
rho				0.207	0.210	0.256	0.275
SE clustered		Yes	Yes	Yes	Yes	Yes	Yes
Mundlak chi2					24.75		
(p-value)					0.17		
Hausman chi2				7.13		6.34	12.05
(p-value)				0.42		0.85	0.28

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Due to the small sample size instead of state dummies only east/west dummy was employed

Table 3: Robustness check, regression results with ability instead of school grades

		Model 1	Model 2
		Basic models	
		OLS	Logit *(av.marg.eff)*
	$success$	0.077** (0.034)	0.080** (0.036)
	$ability$	0.066** (0.033)	0.065** (0.031)
	$status$	0.078 (0.068)	0.074 (0.066)
Π	$male$	-0.016 (0.059)	0.002 (0.056)
	$birth$	-0.106*** (0.030)	-0.120*** (0.030)
	$gymn$	0.213*** (0.066)	0.251*** (0.074)
	abi_age	0.061** (0.029)	0.074*** (0.027)
	$year$	No	No
Ψ	ln_net_hinc	-0.042 (0.053)	-0.034 (0.050)
	$rural$	-0.213*** (0.066)	-0.192*** (0.062)
	mig	0.026 (0.090)	0.049 (0.087)
	pos_sib	0.042 (0.038)	0.033 (0.041)
	$children$	0.047 (0.034)	0.048 (0.034)
	$state$	No	No
Constant		Yes	Yes
Observations		241	241
R-squared		0.251	
F-stat		11,55***	
chi2			53.32***
Log likelihood			-124.1
rho			
SE clustered		Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4: IV-regression results

		Model 1	Model 2	Model 3	Model 4	Model 5
Dependent var.		*success*	*uni*			
Type of model		OLS First stage	2SLS	IV-Probit *(av.marg.eff.)*	Probit *(av.marg.eff.)*	Probit *(av.marg.eff.)*
	success		0.061** (0.031)	0.063** (0.030)		0.078*** (0.023)
Π	*grade*	0.214*** (0.022)	0.080*** (0.017)	0.079*** (0.017)		
	status	-0.030 (0.048)	0.072** (0.033)	0.069** (0.032)		
	male	0.054 (0.043)	0.027 (0.029)	0.025 (0.029)		
	birth	-0.123** (0.061)	0.007 (0.041)	0.008 (0.040)		
	gymn	0.065 (0.052)	0.190*** (0.035)	0.201*** (0.035)		
	abi_age	-0.043** (0.022)	0.095*** (0.015)	0.096*** (0.014)		
	year	Yes	Yes	Yes		
Ψ	*ln_net_hinc*	0.057 (0.051)	0.030 (0.035)	0.036 (0.035)		
	rural	-0.087 (0.071)	-0.052 (0.049)	-0.050 (0.047)		
	mig	-0.211*** (0.060)	0.061 (0.041)	0.061 (0.041)		
	pos_sib	-0.007 (0.028)	0.021 (0.019)	0.017 (0.019)		
	children	-0.028 (0.023)	-0.002 (0.016)	0.000 (0.016)		
	state	Yes	Yes	Yes	Yes	Yes
p_think		0.516*** (0.023)			0.057*** (0.018)	0.014 (0.022)
Constant		Yes	Yes	Yes	Yes	Yes
Observations		992	992	992	992	992
R-squared		0.401			0.008	0.017
F-stat		19.43***				
chi2			273.9***	194.6***	10.06***	20.52***
chi2, exogeneity				0.968		

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

P_think: positive thinking in life

APPENDIX

Figure A-1: The connection between grade and success, by type of school at age 17 – locally weighted regression smoothing

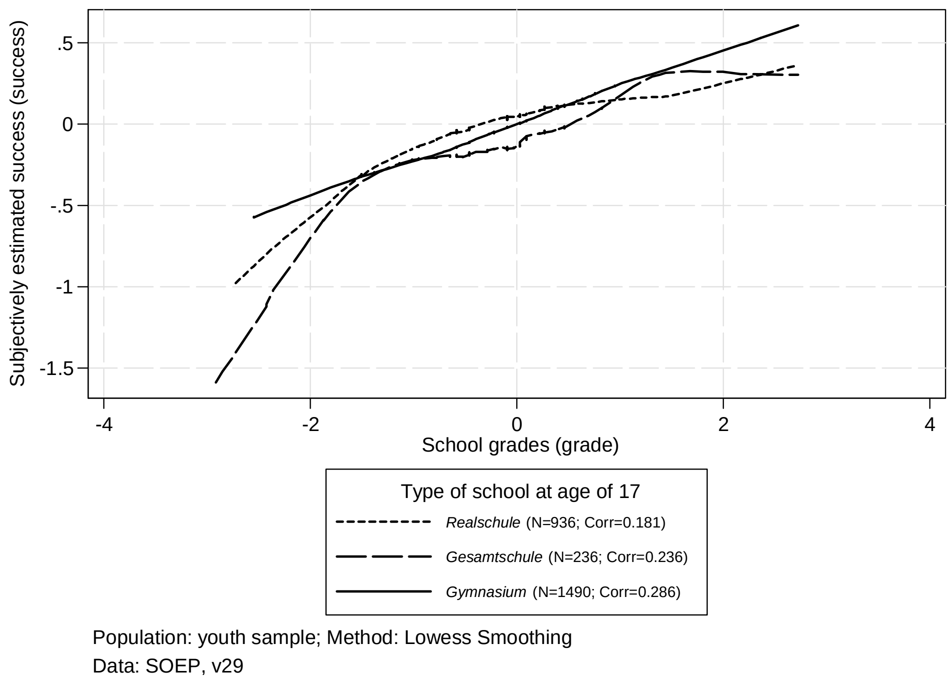


Figure A-2: The connection between plans to apply to university and success, by those with Abitur and without Abitur – locally weighted regression smoothing

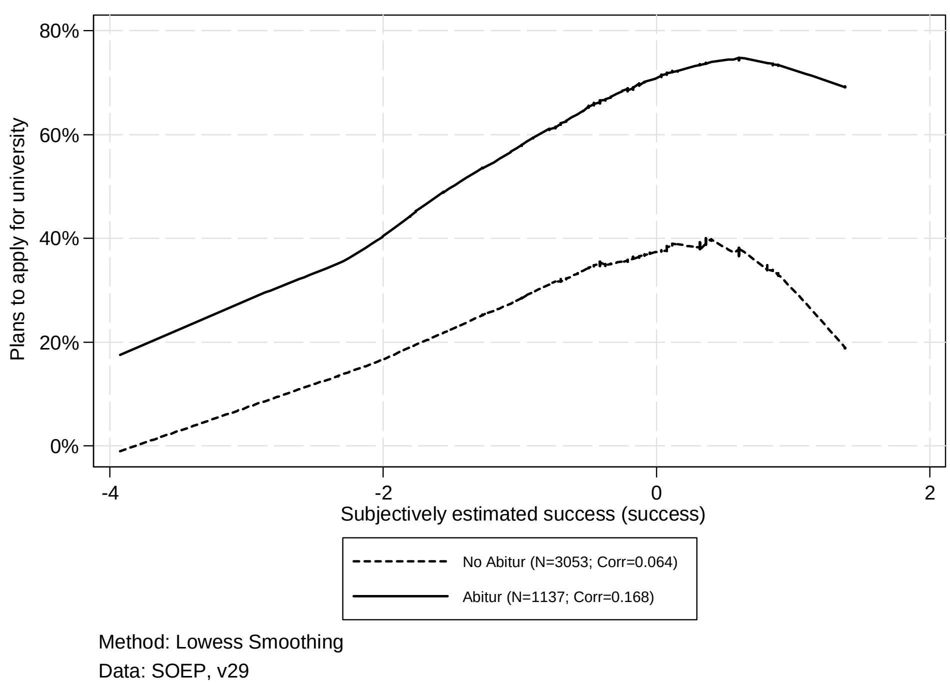


Table A-1: Extracted principal components in the analysis: success; grade; p_think; ability

success	***Subjectively estimated success***	
	Name of the variable used	Correlation with the first factor
	Probability of receiving training/uni slot	0.87
	Probability of finishing training/uni slot	0.87
	Eigenvalue of the first factor	*1.53*
	N	*3881*
grade	***School grades***	
	Name of the variable used	Correlation with the first factor
	Grade in German	0.82
	Grade in mathematics	0.67
	Grade in first foreign language	0.82
	Eigenvalue of the first factor	*1.78*
	N	*3790*
p_think	***Positive thinking***	
	Name of the variable used	Correlation with the first factor
	Probability of finding employment	0.74
	Probability of job success	0.77
	Probability of being self-employed	0.33
	Probability of marriage	0.65
	Probability of several children	0.59
	Eigenvalue of the first factor	*2.03*
	N	*3799*
ability	***Ability***	
	Name of the variable used	Correlation with the first factor
	Verbal test	0.79
	Numerical test	0.74
	Figural test	0.80
	Eigenvalue of the first factor	*1.82*
	N	*1381*

Table A-2: Heckman regression, first stage

		Basic models
		Model 1
		OLS
recc_gymn		0.050*
		(0.028)
gymn_norec		0.079**
		(0.037)
sib_abi	_no	Ref.
	_yes	0.088***
		(0.022)
	_no sibling	-0.043**
		(0.018)
	success	-0.023***
		(0.008)
	grade	0.027***
		(0.008)
	status	-0.001
		(0.017)
Π	male	-0.022
		(0.015)
	birth	0.008
		(0.022)
	gymn	0.324***
		(0.026)
	year	Yes
Ψ	ln_net_hinc	0.018
		(0.016)
	rural	-0.012
		(0.024)
	mig	-0.002
		(0.020)
	pos_sib	-0.019**
		(0.009)
	children	-0.010
		(0.009)
	state	Yes
Constant		Yes
Observations		2,921
R-squared		0.273
chi2		1034.97
Log likelihood		-1374.73

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A-3: Heckman regression, second stage

		Basic models		Family models		Sibling models	
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
		OLS	Binary Logit	Random effect Logit	Mundlak's Random eff.	Sibling random effect Logit	Sibling+Abi Rand. eff. Logit
	$success$	0.038** (0.018)	0.044** (0.018)	0.047** (0.021)	0.113** (0.053)		
	$\Delta success$					0.066* (0.037)	0.144** (0.063)
	$\overline{success}$					0.073** (0.037)	0.145** (0.064)
	$grade$	0.083*** (0.018)	0.079*** (0.017)	0.085*** (0.019)	0.038 (0.038)		
	$\Delta grade$					0.034 (0.032)	0.032 (0.043)
	$\overline{grade}$					0.111*** (0.034)	0.101** (0.047)
	$status$	0.082** (0.033)	0.081** (0.032)	0.086** (0.035)	-0.062 (0.151)	0.120** (0.048)	0.070 (0.080)
Π	$male$	0.039 (0.030)	0.046 (0.030)	0.050 (0.032)	0.036 (0.071)	0.018 (0.043)	0.029 (0.065)
	$birth$	0.004 (0.044)	0.005 (0.042)	0.012 (0.044)	0.075 (0.093)	0.088 (0.063)	0.002 (0.018)
	$gymn$	0.108 (0.124)	0.065 (0.124)	0.077 (0.134)	0.271* (0.158)	0.328 (0.203)	0.185 (0.233)
	abi_age	0.092*** (0.014)	0.095*** (0.014)	0.102*** (0.016)	0.051 (0.042)	0.077*** (0.022)	0.105*** (0.035)
	$year$	Yes	Yes	Yes	Yes	Yes	Yes
Ψ	ln_net_hinc	0.031 (0.036)	0.030 (0.035)	0.034 (0.038)	0.070 (0.171)	0.020 (0.054)	0.001 (0.089)
	$rural$	-0.055 (0.050)	-0.048 (0.047)	-0.053 (0.052)	-0.447 (0.472)	-0.128* (0.075)	-0.027 (0.085)
	mig	0.049 (0.041)	0.052 (0.042)	0.058 (0.046)	0.054 (0.042)	0.090 (0.061)	0.034 (0.111)
	pos_sib	0.022 (0.020)	0.020 (0.019)	0.023 (0.021)	0.147 (0.091)	-0.011 (0.027)	0.128** (0.057)
	$children$	-0.001 (0.018)	-0.002 (0.018)	-0.002 (0.018)	0.051 (0.112)	0.004 (0.024)	0.047 (0.060)
	$state$	Yes	Yes	Yes	Yes	Yes	Yes#
Mean of indep. var.					Yes		
$mills\_ratio$		-0.093 (0.124)	-0.158 (0.131)	-0.166 (0.141)	-0.077 (0.132)	0.054 (0.216)	-0.212 (0.264)
Constant		Yes	Yes	Yes	Yes	Yes	Yes
Observations		1,021	1,021	1,021	1,021	604	232
No. of families		865	865	865	865	448	111
R-squared		0.224					
F-stat		13.18***					
chi2			172.4***	71.08***	295.6***	50.02	26.80**
Log likelihood			-573.6	-572.5		-334.6	-125.0
rho				0.197	0.211	0.258	0.295
SE clustered		Yes	Yes	Yes	Yes	Yes	Yes
Mundlak chi2 (p-value)					24.69 0.17		
Hausman chi2 (p-value)				7.34 0.39		5.60 0.59	9.14 0.61

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table A-4: Variables in SOEP used to construct variables in the analysis

		Data file	Variables used	Definition
Dependent variables				
	uni *uni2*	bioedu	- bet7obs - bet7year - bet5obs - bet5year - bet4type - bex4cert	- first entry into tertiary education, observed - first entry into tertiary education, year - first entry into vocational training, observed - first entry into vocational training, year first exit from school, type of school - highest school-leaving certificate obtained
	abi	bioedu	- bet4type - bex4cert	- first exit from school, type of school - highest school-leaving certificate obtained
Independent variables				
	Probability of success, personal estimations			
	success	bioage17	- bywaausp - bywaerfa	- Did you receive training or a university place in your preferred field? - Did you successfully finish your training or university studies?
	Probability of success, derived from school performance			
	grade	bioage17	- byntdeut - byntmath - byntfmd1	- School grade: German, - School grade: mathematics - School grade: first foreign language
	Social status			
	status	\$pgen \$pequiv	egp d11108[18]	Erikson, Goldthorpe and Portocarero (1979) class scheme Education with respect to high school
Π	*Vector of individual-specific variables*			
	male	ppfad	sex	Gender
	birth	bioage17	bygebjah	Year of birth
	gymn	bioage17	- byschbes - byschabs	- Type of school visited currently - Type of certificate
	abi_age	bioedu	- bet4year - bex4age	- First exit from school, year - Last observed year in school
	year	bioage17	erhebj	Year of survey
Ψ	*Vector of family-specific variables*			
	ln_net_hinc	\$pequiv	i11102	Net (post-government) household income
	state	\$pequiv	l11101	State of residence
	rural	\$hbrutto	regtyp	Spatial category by BBSR[19]
	mig	ppfad	migback	Migration background
	pos_sib	biosib	pos_sib	Position among siblings
	children	\$pequiv	h11101 h11102	Number of household members age 0–14 Number of household members age 15–18
H	*Variables used in the first stage of Heckman regression*			
	recc_gymn	bioage17	byempfeh	School recommendation after elementary school
	gymn_norec	bioage17	- byempfeh - byschbes - byschabs	- School recommendation after elementary school - Type of school visited currently - Type of certificate
	sib_abi	bioedu	- bet4type - bex4cert - bymnr - byvnr	- First exit from school, type of school - Highest school-leaving certificate obtained - Mother ID - Father ID
abil	*Proxy for ability*			
	abil	cogdj	- analog - rechenz - matrize	- Number of correct answers on a verbal test (analogies) - Number of correct answers on a numerical test (arithmetic) - Number of correct answers on a figural test (figures)

[18] Employed only in robustness check.

[19] The Federal Institute for Research on Building, Urban Affairs and Spatial Development, http://www.bbsr.bund.de/

Big Five personality measures			
op	$page17	$j9104 $j9109 $j9114 $j9116	Openness to experience
co	$page17	$j9101 $j9107 $j9111	Conscientiousness
ex	$page17	$j9102 $j9108 $j9112	Extraversion
ag	$page17	$j9103 $j9106 $j9113	Agreeableness
ne	$page17	$j9105 $j9110 $j9115	Neuroticism
Rotter's locus of control			
ro_ex	$page17	$j9003 $j9004 $j9006 $j9009	External control
ro_in	$page17	$j9001 $j9005 $j9008	Internal control
Positive thinking			
p_think	bioage17	- bywaarbp - bywaberf - bywaselb - bywaheir - bywakidm	- Probability of finding employment - Probability of job success - Probability of being self-employed - Probability of marriage - Probability of several children