Working Paper No. 120

Subjective completion beliefs and the demand for post-secondary education

Johannes S. Kunz and Kevin E. Staub

Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

UNIVERSITÄT BERN

Working Paper No. 120

Subjective completion beliefs and the demand for post-secondary education

Johannes S. Kunz and Kevin E. Staub

November 2016

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

Subjective completion beliefs and the demand for post-secondary education*

Johannes S. Kunz
University of Zurich

Kevin E. Staub
University of Melbourne and IZA

November 1, 2016

Abstract

The outcome of pursuing an upper or post-secondary education degree is uncertain. A student might not complete a chosen degree for a number of reasons, such as insufficient academic preparation or financial constraints. Thus, when considering whether to invest in post-secondary education, students must factor their probability of completing the degree into their decision. We study the role of this uncertainty in education choices using representative survey data from Germany. Students' subjective beliefs about the probability of completing a post-secondary education were elicited prior to them finishing their secondary education. We relate these subjective completion probabilities to students' subsequent education choices and outcomes. We find that these early beliefs are predictive of intentions to invest in education, actual subsequent investments in education, and degree completion. A structural choice model of sequential investment further reveals that the association between completion beliefs and investment choices is strongest for students with low academic skills and low preferences for post-secondary education.

Keywords: Subjective beliefs; Subjective probabilities; Completion uncertainty; Post-secondary education; Human capital investment.
JEL classification: I21, I26, J24

*Addresses for correspondence: J. Kunz (corresponding author), University of Zürich, Department of Economics, Zürichbergstrasse 14, CH-8039 Zürich; johannes.kunz@econ.uzh.ch. K. Staub, Department of Economics, 111 Barry Street, The University of Melbourne, 3010 VIC, Australia; kevin.staub@unimelb.edu.au.
We thank Joseph Altonji, Orazio Attanasio, Uschi Backes-Gellner, Dan Black, Timo Boppart, Gregory Crawford, Thomas Dohmen, Ernst Fehr, Joe Hirschberg, Jamie Gloor, Uli Kaiser, Ed Lazear, Edwin Leuven, Jenny Lye, EeCheng Ong, Philippe Ruh, Florian Schaffner, Carmit Segal, Andreas Steinhauer, Steven Stillman, Sabrina Studer, Rainer Winkelmann, Joachim Winter, and seminar participants at Adelaide, EEA (Geneva), ESAM (Sydney), ESPE (Berlin), Melbourne, the Ski and Labor Seminar, SSES (Lugano), the Workshop of Personnel Economics and Economics of Education, Zurich Workshop on Economics (Schaffhausen), and Zurich, for helpful comments and suggestions; and Daniel Auer and Luca Tonizzo for very helpful research assistance.

1 Introduction

After finishing compulsory education, students have to decide whether to continue investing in education. This choice is one of the most important career decisions young adults have to make, and is made under partial information; beliefs held at this early stage are thus central to post-secondary education choices. In this paper, we focus on one major source of uncertainty: the probability of dropping out of a chosen education track. We incorporate a measure of the perceived or subjective probability of being able to complete a post-secondary degree into simple models of education choice, using data from a large representative survey from Germany. This data offers at least two characteristics that contribute to our understanding of young adults' education decisions. First, subjective completion beliefs are elicited shortly *before* students finish their compulsory secondary education; the data track students over the next 14 years, which allows us to observe their subsequent education outcomes. In contrast, a large part of the literature focuses on beliefs of students who are already enrolled in college. At that point in time, students' beliefs are likely to have been updated by new information, and are thus likely to be different from the original beliefs upon which investment decisions were based. Second, the representative population sample allows us to extend the previous literature, which has been confined to college education, to more vocational forms of post-secondary education, such as different types of apprenticeships. Since students from less advantaged economic backgrounds are more likely to invest in vocational education, it is necessary to take these education tracks into account when designing informational policy interventions.

We have three key objectives. First, we study the determinants of these subjective completion beliefs. Specifically, we assess the relative importance that students assign to factors such as past academic performance, personality measures, risk attitudes, family background, and labor market characteristics. Second, we quantify the importance of subjective completion beliefs for three key outcomes: the intention to invest in education, the actual subsequent investment, and the actual completion of the chosen education track. In addition to estimating the persistence of these early beliefs as predictors of education outcomes, we can compare the actual influence of academic, personality, risk, family, and local labor and education market variables on degree completion vis-à-vis belief formation. Third, we delve deeper into the relationship between beliefs and education by incorporating subjective probabilities into a sequential model of human capital investment. The model combines the different post-secondary options of the German system with forward-looking behavior and unobserved tastes or preferences for education that are correlated across choices.

Our findings suggest that the most important determinants in the formation of completion beliefs are students' academic ability and personality. In contrast, subjective risk attitudes, family background variables (such as household income or parents' education), and local labor market variables (such as youth unemployment or various indicators of educational demand and supply) are given comparatively less weight by students when they are about to finish secondary education. These determinants have similar effects on intentions to invest in education. However, actual investment and completion depend to a larger extent on family characteristics, the state of the local labor market, and regional supply and demand in the post-secondary education market. When we consider different post-secondary options

separately, we find that academic ability is the main driver of subjective beliefs for those choosing a university education. In contrast, the subjective beliefs of those choosing a vocational education seem to be driven more by non-academic factors. We also find that beliefs are most decisive for youths with low academic ability and weak preferences for education, a group that has largely been ignored in the present literature.

Our results point to a strong persistence of early completion beliefs in post-secondary education choices and outcomes; beliefs are statistically significant predictors of intentions to invest, investments, and completion, even after accounting for a large set of background variables. The effects are also substantial in an economic sense. For instance, our models predict that an increase in completion beliefs by one standard deviation would be associated with a reduction in the number of youths who fail to invest in post-secondary education by about 20 to 30 percent, depending on the model specification. Although these estimates are correlational in nature, we show that they are robust in a bounding analysis which takes potential selection on unobservables into account.

Understanding the role of uncertainty in individuals' post-secondary education choices is essential for designing effective education policies. For instance, if students' expectations are misaligned, providing additional information can be a cost-effective measure to enhance education choices and eventual career success (e.g., Bettinger et al., 2012; Jensen, 2010).

1.1 Related literature

The incorporation of completion probabilities into theories of education choice has a long standing (e.g., Altonji 1993; Comay, Melnik and Pollatschek 1973; Manski 1989; see also Bound and Turner 2011). This literature emphasizes that "[d]ifferences in dropout probabilities may be more important than differences in *ex post* payoffs in determining the *ex ante* return to attending a particular school" (Altonji, 1993, p74). Empirical approaches based on structural assumptions that distinguish *ex ante* from *ex post* returns to education include Carneiro et al. (2003), Cunha, Heckman and Navarro (2005), Cunha and Heckman (2007), and Foley, Gallipoli and Green (2014).[1] However, there is relatively little empirical work using completion probabilities directly; notable exceptions using predicted college completion probabilities include Hussey and Swinton (2011), Fossen and Glocker (2014), and Castex (2015). In general, when studying choice under uncertainty, researchers have to make assumptions about how expectations are formed (Manski, 2004), and, most commonly, such work relies on rational expectations, e.g., that individuals' predictions are unbiased. To avoid imposing rational expectations, the most widely used alternative, which we follow in this paper, is to apply direct measures of elicited

[1] Completion uncertainty also has important consequences beyond aggravating wage uncertainty. For example, various non-pecuniary aspects have been shown to be relevant to educational choice (see Oreopoulos and Salvanes, 2011, for a recent summary). In order to benefit from them, staying in the chosen educational path and/or completing the degree might be crucial. Studies using elicited subjective beliefs about labor market prospects consistently find the (non-financial) consumption value of education or major-specific unobserved tastes to be the main drivers of educational choices (i.e., Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*). Similar results are found in structural approaches that do not use subjective beliefs. For instance, D'Haultfoeuille and Maurel (2013) use a sophisticated Roy model and find non-pecuniary aspects to be predominant in educational choice. Such preference-related factors are not affected by pure labor market uncertainty, but they can be affected by completion uncertainty.

subjective beliefs (Manski, 2004). Several studies have shown subjective beliefs to be meaningful in education choice models, and often superior to those constructed by rational expectations models (e.g., Attanasio and Kaufmann, 2014; Huntington-Klein, 2015*b*; Stinebrickner and Stinebrickner, 2012; Zafar, 2011*a*) and a growing literature on college major choice uses such subjective beliefs (e.g., Arcidiacono et al., 2014; Arcidiacono, Hotz and Kang, 2013; Hastings et al., 2016; Huntington-Klein, 2016; Stinebrickner and Stinebrickner, 2012, 2014*a*,*b*; Wiswall and Zafar, 2015*a*). In contrast to our paper, most of these studies assess beliefs about returns to or costs of educational choices, rather than about the likelihood of completion. In this respect, our analysis is most closely related to Wiswall and Zafar (2015*a*), who also use students' subjective completion beliefs. We contribute to this literature in various dimensions: Our paper is the first to study subjective completion beliefs formed before the end of secondary education in a population survey; we follow students over time until they complete their post-secondary education, and we integrate elicited subjective completion probabilities into a sequential model of educational choice, as motivated by the theoretical literature. Most of the existing studies on the demand for post-secondary education focus solely on investment, rather than on intentions or completion;[2] we add to the literature by providing evidence on the relationship between subjective completion beliefs and each of these three outcomes in a unifying framework.

Besides beliefs about returns to education, beliefs about one's own academic ability have recently been highlighted as a key factor in college choice (Arcidiacono, Hotz and Kang, 2013; Bond et al., 2016; Bulman, 2015; Stange, 2012; Stinebrickner and Stinebrickner, 2012, 2014*b*; Zafar, 2011*b*): students learn about their own ability by observing their grade point averages [GPA], and this updated information on their own ability is a main determinant of college enrollment, college-major choice, and dropout decisions.[3] While belief updating is essential for staying in school, results in this literature also point towards preexisting beliefs as key determinants of later investment and success in education. For example, Bond et al. (2016, p2) find that belief updating in response to SAT scores is too modest to explain the variation in college application choices, and conclude that there is a substantial amount of "inertia" in college choices, in the sense that they are "predetermined by non-academic factors and preexisting beliefs". In a similar vein, Stinebrickner and Stinebrickner (2014*a*, p468) stress the importance of timing interventions to inform students about their own ability before college entrance. We add to the literature by studying the persistence in education choice due to the ex ante perceived probability of completing a post-secondary education, using a representative population survey.[4] Our evidence supports and extends Zafar's presumption that "prior belief[s] [at the start of college] continue[s] to be important. In attempting to understand the choice of college majors, it might be useful to focus on students at earlier stages of their schooling (for example, in high school) and analyze their subjective beliefs" (Zafar, 2011*b*, p339f).

[2]Notable exceptions are Turner (2004), Venti and Wise (1983) and Light and Strayer (2000) for completion. Similarly, the literature on intentions is still comparatively small, although it has been growing recently (e.g., Christofides et al., 2015; Wiswall and Zafar, 2015*b*; Zachary and Zafar, 2015).

[3]For evidence on students applying for college, see Bond et al. (2016); Bulman (2015), and for students enrolled in college, Stinebrickner and Stinebrickner (2012, 2014*a*,*b*); Zafar (2011*b*).

[4]Thus far, studies have focused mainly on single institutions rather than representative samples. Exceptions include Milla (2014) and Bond et al. (2016). Both assess students enrolled in college or who applied for college in representative samples. As opposed to us, they do not use an elicited measure of subjective beliefs, and focus on the selected population applying to, or enrolled in, college.

Our study is also closely related to the literature on non-cognitive determinants of education success (see, Almlund et al., 2011; Borghans et al., 2006). In particular, we show how subjective completion beliefs relate to the Big Five personality measures, risk attitudes, and locus of control, all of which are now ubiquitous in economic applications (see, for example, Borghans et al., 2006; Caliendo, Cobb-Clark and Uhlendorff, 2015; Dohmen et al., 2010). Moreover, we use estimates from our structural model to decompose the effects of academic ability, locus of control, and parental background into their direct impact on investment and their indirect impact via subjective completion beliefs. Of special interest to our design is the locus of control. Coleman and DeLeire (2003) hypothesize that students with a more internal locus of control (i.e., students who believe their actions affect their outcomes) have higher subjective beliefs about their own returns to education, which leads them to exert more effort and to invest in their human capital (for a recent review on the locus of control, see Cobb-Clark, 2015). Our results support the hypothesis that one's locus of control affects education choices via subjective beliefs: We estimate that virtually the entire effect of locus of control acts through subjective beliefs (for other variables, subjective beliefs play only a partial role). Thus, this paper also adds to recent contributions assessing the role of subjective beliefs as a mediator and a potential explanation of educational differentials in parental unemployment (Pinger, 2015), family background (Keller and Neidhöfer, 2014), or gender and migration (Tolsma, Need and De Jong, 2010). Our framework might prove useful in studying the mediating role of subjective beliefs, since it integrates investment in upper and post-secondary education jointly in both reduced-form and structural models.

1.2 Outline of the paper

The remainder of this study proceeds as follows: In Section 2, we describe the institutional features of the education system in Germany and present the data. In Section 3, we assess determinants of subjective completion beliefs, relate these beliefs to education outcomes, and explain how the impact of subjective beliefs varies with selection on observables and unobservables. We develop and estimate a structural model of sequential human capital investment in Section 4. Section 5 concludes our paper by discussing and reviewing our key findings.

2 Institutional setting, data, and descriptive statistics

2.1 Institutional setting

A simplified version of Germany's education system is depicted in Figure 1, in which we briefly summarize the system's key features that are relevant to our analysis (more information can be found in Wölfel and Heineck, 2012). The German education system is characterized by early tracking, which takes place after grade four (elementary school), at age nine to eleven years.[5] Based on grades and teachers' recommendations, the children are tracked into three streams according to their academic

[5] With the exceptions of Berlin and Brandenburg, which track after grade six (ages 11 to 13 years).

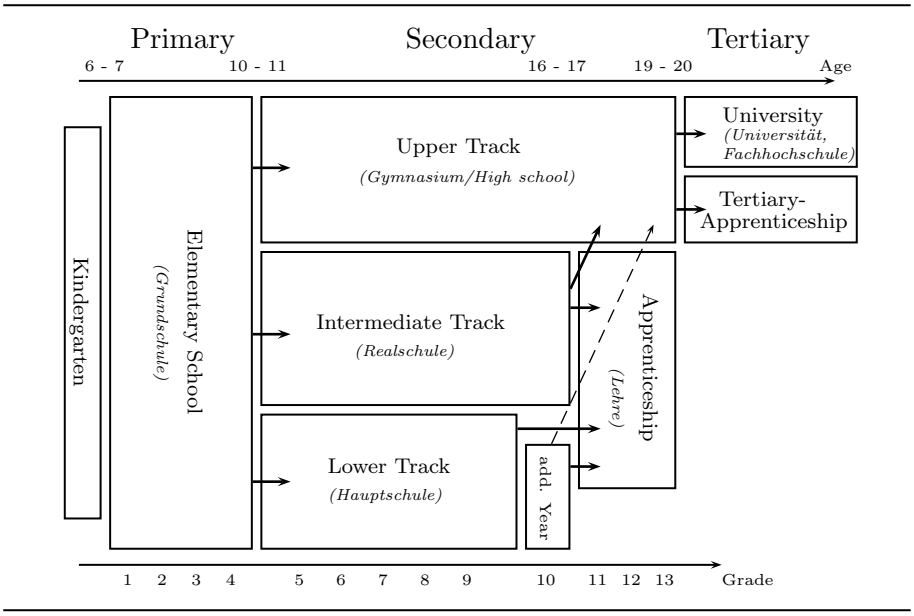


Figure 1: The German education system

Source: Adaptation and extension of the overview provided by Wölfel and Heineck (2012).

ability.[6] The statistical agency in Germany (Statistisches Bundesamt, 2014, p27) reported that in 2012, ten percent of children were assigned to lower-track schools, 19 percent to intermediate-track schools, and 40 percent to upper-track schools (high schools). The remaining children enrolled in other, so-called comprehensive schools; in these, children are not separated, but follow the same tracking into upper, intermediate and lower streams at a later stage.

At the time of entering the survey population, youths—ages 16 to 17 years—are in the midst of deciding whether to enroll in post-secondary education. Students completing lower or intermediary tracks have the opportunity to apply for and start a profession-specific apprenticeship or a vocational education.[7] Although investing directly in an apprenticeship is the most common path, they could also enroll in a consecutive school-track that leads to the *university entrance qualification* (German: *Abitur*), the equivalent of a high school degree.[8] This high school degree is valuable even if students do not plan to attend university; students with high school degrees typically have better chances, compared to their peers who completed a lower track, when applying for highly competitive apprenticeship positions. In fact, some apprenticeship positions are exclusively available to high school graduates. In 2010, 20.9 percent of the newly signed apprenticeship contracts went to students with high school degrees (Statistisches Bundesamt, 2011, p1004). Thus, we model this path separately and refer to it as *tertiary apprenticeship*.

The decision is different for students already enrolled in high school (the upper track). In principle, they could also start an apprenticeship straight away. Alternatively, they could continue with their high school education and, upon completion of high school, either enroll in university or start a tertiary

[6]The binding nature of these recommendations varies across states.

[7]Students who started an apprenticeship before entering the survey population are excluded from our analyses. However, in 2011, only 10.6 percent started an apprenticeship before the age of 17 years (Statistisches Bundesamt, 2013, p17).

[8]Due to the limited time horizon of our sample we focus on early investment. The possibility of visiting complementary courses that allow students to go to university after apprenticeship completion is not modeled separately.

apprenticeship.[9] Yet, their default choice is to continue in high school, as they are already enrolled and do not have to make an active choice to enroll.[10]

For instance, of the class of 2011, four percent dropped out without a secondary degree, 17 percent completed the lower track, 36 percent the intermediary track, and 43 percent obtained high school degrees (Statistisches Bundesamt, 2013, p7). It is important to realize that in Germany, an apprenticeship degree has a high standing and can have a reputation similar to a university degree—especially if the apprenticeship degree is acquired after completing high school.

2.2 Data sources

Our primary data source is the German Socio-Economic Panel [SOEP], a large-scale representative household panel data set. We focus on youths, ages 16 to 17 years, who have newly entered the survey population by answering the youth questionnaire between 2000 and 2013. The SOEP is a yearly household panel that provides a rich set of parental background information, and follows the young adults over time up to 14 years. Additionally, we combine the individual-level data with regional labor market information and educational supply and demand measures based on 96 geographic regions, which we will refer to as Ror (for their German name *Raumordnungsregionen*).[11] All regional information is matched according to the individual's residency when answering the youth questionnaire, and lagged by one year to avoid endogeneity or reverse causality. We only use variables assessed in the youth questionnaire as covariates to avoid any biases from conditioning on (future) outcomes.[12]

2.3 Outcome variables

We are interested in a student's intention to invest in post-secondary education, actual investment, and actual completion. We distinguish between not investing, and investing in one of the three most commonly taken educational tracks in Germany: At the time of answering the youth questionnaire, youths can choose not to invest ($j = 0$), to invest in an apprenticeship directly after completing either the lower or intermediate school track ($j = 1$), or to continue high school. After completing high school, students can decide whether to invest in an apprenticeship ($j = 2$) or in a university degree ($j = 3$).

To assess their post-secondary investment intention, we use a self-reported measure elicited in the youth questionnaire: Students are asked to indicate which further educational degree, if any, they

[9] Additionally, students could also drop out after completing high school, but this rarely occurs in practice (see also Fossen and Glocker, 2014). Note that here university subsumes universities of applied sciences. While it would be interesting to consider those separately, we have to leave this to future research due to our current sample size.

[10] All regressions include the indicator variable "In high school with 17". We also estimated the regressions of our main Table 3 separately for the two groups and present the results in Appendix, Table D4.

[11] A map of Ror is provided in Appendix, Figure B1. The data source is INKAR 2013 provided by the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR, 2013). For more information, see Pinger (2015) who also uses this additional data source. Moreover, we add the number of universities (higher learning institutions) as a proxy for distance to university provided by the statistical agency of Germany (Statistisches Bundesamt).

[12] One of the motivations for the structural model presented in Section 4 is to account for sequential decision-making and to avoid associated biases.

plan to complete. The answers correspond to $j = 0, 1, 2, 3$.[13] To assess actual post-secondary educational investment, we use longitudinal information from the SOEP two years after the individual has answered the youth questionnaire. If, after this period, the individual is observed to have started an apprenticeship, a tertiary apprenticeship, or a university degree, this investment is recorded accordingly; otherwise, we record the individual as not having invested in post-secondary education. Actual post-secondary completion is assessed five years after the individual has answered the youth questionnaire, and constructed analogously to investment intention and actual investment.

2.4 Subjective completion beliefs

Our main variable of interest is subjective completion belief, p_i, that is assessed by the following question in the youth questionnaire:

> Think about your future in your job and private life: how probable is it, in your opinion, that the following events will occur?
>
> [Please check off a probability on the scale from 0 percent to 100 percent.]
>
> *You successfully finish your vocational training or university studies?*[14]

When we incorporate this measure into models of education choice in the next sections, we interpret this belief as the subjective probability of graduation from a post-secondary education track conditional on attendance, i.e., $p_i \equiv Pr_i(complete = 1|invest = 1)$ for student i.

Another interpretation would be that p_i elicits the unconditional completion probability entailing also the belief of investing: $Pr_i(complete = 1) = Pr_i(invest = 1)Pr_i(complete = 1|invest = 1)$. However, this is unlikely for at least two reasons. First, the question immediately preceding the subjective completion belief question in the survey directly assesses students' probability of investment in their preferred track, $Pr_i(invest = 1)$, which raises our confidence that students were thinking about p_i as the conditional probability. Second, the two subjective beliefs questions provide bounds on each other that can be used to check their coherency. Using the self-reported $Pr_i(invest = 1)$, if p_i referred to $Pr_i(complete = 1)$, it could not be larger than $Pr_i(invest = 1)$. However, this interpretation of p_i as the unconditional probability is strongly rejected by the data.[15] In the appendix, we present results where we interpret p_i as the unconditional probability, in which case our models need to be modified to control for intentions and the subjective investment probability; our findings are robust to this interpretation (cf. Table D2).

[13] That means that, for consistency with the other outcome variables, students who want to enroll in an apprenticeship first and then continue with supplementary courses that prepare for university are subsumed into the apprenticeship category.

[14] Students could answer on an eleven point scale: 0%, 10%, 20%,..., 100%. The exact wording in German is: *Wenn Sie sich einmal Ihre berufliche und private Zukunft vorstellen: Wie wahrscheinlich ist es, dass die folgenden Entwicklungen eintreten werden?* [Stufen Sie bitte jeweils die Wahrscheinlichkeit auf einer Skala ein, die von 0 Prozent bis 100 Prozent geht.] *Ihre Ausbildung oder Ihr Studium erfolgreich abschließen?*

[15] We acknowledge that this could also be interpreted as students having difficulties to respond consistently to subjective beliefs questions: their responses would simply include some error. However, the magnitude of the inconsistency is so large that it cannot reasonably be interpreted as errors, and clearly points to p_i being understood as the conditional probability.

2.5 Estimation sample

As stated before, we exclude from the estimation sample all individuals who have already started an apprenticeship at the time of responding to the youth questionnaire. Moreover, we exclude students with missing information in the core variables: subjective belief, education outcomes, and grade point average [GPA]. All other missing information are included along with corresponding indicator variables for missing information. This selection results in a sample size of 3,610 individual observations. In the longitudinal analysis, we additionally require at least two years of information to assess the end of secondary education and the start of a post-secondary education, which reduces the observations to 2,116. To assess educational completion, we restrict the sample to students who responded for at least five years of data collection, resulting in 1,372 observations.[16]

2.6 Descriptive statistics

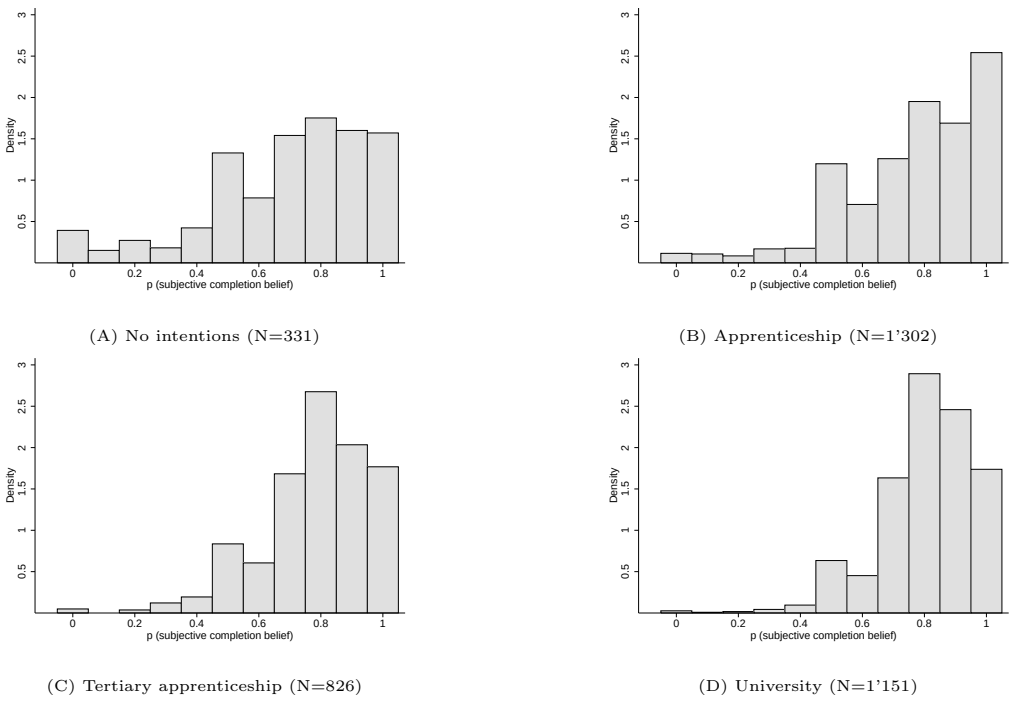
In Figure 2, we plot histograms of subjective beliefs by students' educational intentions. Overall, German students appear to be confident about finishing a post-secondary education, as most report a subjective probability above 50%. The distributions of students intending to invest in paths that first require a high school degree (tertiary apprenticeship and university, Panels C and D) are very similar in shape. Yet, the subjective beliefs of youths who intend to complete a university degree are more concentrated and slightly shifted to the right compared to their high school counterparts who aim for a tertiary apprenticeship position. The mode of the distribution of students who intend to start an apprenticeship without finishing high school lies at 100%. Finally, students with no intention to invest in further education display a much larger variance in their beliefs. The fact that positive beliefs are found among all intended investment levels, including the intention not to invest in further training or education, can be rationalized in a simple expected utility framework where students weight their utility from education by their beliefs about their completion probabilities, and report their intentions based on their highest expected utility. This is also the interpretation we pursue in the following analysis.

Analogously, Table 1 presents descriptive statistics by intended investment level for our baseline sample.[17] At the bottom of the table, we present the sample shares of the intentions to invest: The modal category is the intended completion of an apprenticeship (36 percent), followed by university studies (32 percent). A substantial share wants to pursue a high school education and an apprenticeship (tertiary apprenticeship, 23 percent), and roughly nine percent do not intend to invest in any professional education. It is reassuring that these sample statistics are consistent with the population statistics presented above (Statistisches Bundesamt, 2013).

Individuals who intend to invest in a university education have the highest average confidence level about successfully completing their post-secondary education—80.5 percent. However, all young adults

[16] More information on missing values and the construction of the variables can be found in the Appendix, Table B2.

[17] Unconditional descriptive statistics for the various subsamples considered in the analysis below are presented in the Appendix, Table B1.



(A) No intentions (N=331)

(B) Apprenticeship (N=1'302)

(C) Tertiary apprenticeship (N=826)

(D) University (N=1'151)

Figure 2: Subjective completion beliefs by intentions to invest in post-secondary education
Source: SOEP 2000-2013, own calculations.

who report an intention to invest in some education exhibit a comparable level of completion beliefs, as opposed to those without educational intentions whose average belief at 69 percent is visibly lower. Educational investment intentions are positively related to our measures of academic ability. Average GPA clearly increases with the academic rigour of the intended post-secondary education, increasing from apprenticeship to tertiary apprenticeship to university.[18] Prior track recommendations at the age of ten years seem to be a good indicator for the intentions up to seven years later, which could either be a reflection of effective ability streaming, or a manifestation of students' expectations as a result of early tracking. Interestingly, having no educational intentions occurs in all tracks, and the largest share of students without intentions is found in high school. This could be explained by a default effect, as the survey elicits these intentions at a time when students not enrolled in high school have to make an active decision, as opposed to high school students, who can follow their track and decide after obtaining a high school degree.

We assess youths' personalities by their locus of control, risk attitudes, and the Big Five personality inventory.[19] Educational intentions are positively associated with the locus of control, which measures the degree to which a person believes her life is under her own control. Among the standard Big Five inventory, intentions increase with openness, agreeableness, and extraversion but are less monotonically related to conscientiousness or neuroticism. Intending to go to university is positively associated with

[18]The GPA refers to the student's average of the German and Math grade, which is standardized over the sample population we present in our main results from Table 3. We also standardize the GPA within school track in Appendix, Table D5 to show that the choice of standardization does not drive the results. We further standardize GPA within federal states to show that different grading levels do not affect the results, see Appendix, Table D6.

[19]We standardize all the principal components of the personality variables (all but risk attitudes, which are assessed by one question only). The locus of control has been developed by Rotter (1966), the Big Five inventory by Costa and McCrae (1992) and validated in the SOEP version by Hahn, Gottschling and Spinath (2012). Risk attitudes have been introduced and extensively studied by Dohmen et al. (2011).

Table 1: Descriptive statistics by intentions to invest in post-secondary education

	By intended investment level				Total
Variables	0	1	2	3	
p	0.692	0.768	0.782	0.805	0.776
	(0.258)	(0.220)	(0.176)	(0.155)	(0.198)
Academic variables					
GPA (std)	-0.269	-0.288	-0.000	0.404	0.000
	(1.035)	(0.919)	(0.902)	(1.007)	(1.000)
Rec: Lowest Track (yes/no)	0.166	0.265	0.044	0.034	0.132
	(0.373)	(0.441)	(0.204)	(0.181)	(0.338)
Rec: Intermediate Track (yes/no)	0.248	0.351	0.271	0.150	0.259
	(0.432)	(0.477)	(0.445)	(0.358)	(0.438)
Rec: High school (yes/no)	0.284	0.092	0.524	0.698	0.402
	(0.452)	(0.289)	(0.500)	(0.459)	(0.490)
In high school (yes/no)	0.272	0.015	0.596	0.767	0.411
	(0.446)	(0.123)	(0.491)	(0.423)	(0.492)
Personality variables					
Locus of control (std)	-0.183	-0.155	0.028	0.209	0.000
	(1.066)	(0.961)	(0.911)	(0.894)	(0.952)
Risk attitudes (std)	-0.053	-0.012	-0.035	0.054	0.000
	(0.983)	(0.986)	(0.935)	(0.929)	(0.957)
Openness (std)	-0.180	-0.136	0.048	0.171	0.000
	(1.012)	(0.957)	(0.879)	(0.928)	(0.946)
Agreeableness (std)	-0.110	-0.049	0.067	0.039	-0.000
	(0.929)	(0.991)	(0.909)	(0.922)	(0.947)
Extraversion (std)	-0.127	-0.050	0.021	0.079	0.000
	(0.954)	(0.915)	(0.919)	(0.991)	(0.946)
Neuroticism (std)	0.068	0.010	0.052	-0.068	-0.000
	(0.960)	(0.925)	(0.886)	(1.003)	(0.946)
Conscientiousness (std)	-0.145	0.051	-0.059	0.026	0.000
	(0.985)	(0.943)	(0.915)	(0.954)	(0.946)
Family background variables					
Female (yes/no)	0.486	0.454	0.541	0.539	0.504
	(0.501)	(0.498)	(0.499)	(0.499)	(0.500)
No. siblings	1.613	1.710	1.433	1.496	1.570
	(1.461)	(1.494)	(1.206)	(1.099)	(1.316)
Second-generation migrant (yes/no)	0.746	0.680	0.574	0.557	0.623
	(0.436)	(0.466)	(0.495)	(0.497)	(0.485)
Parent college-educated (yes/no)	0.199	0.101	0.306	0.495	0.283
	(0.400)	(0.301)	(0.461)	(0.500)	(0.450)
Parent cur. unemployed (yes/no)	0.124	0.160	0.087	0.045	0.103
	(0.330)	(0.367)	(0.282)	(0.208)	(0.304)
Log. net household income	10.019	9.890	10.624	10.855	10.377
	(2.231)	(2.216)	(1.358)	(1.295)	(1.834)
Local labor market variables					
Cyclical youth unemployment (in Ror)	0.154	0.101	0.043	0.041	0.074
	(1.079)	(1.044)	(0.982)	(1.020)	(1.026)
No. of apprenticeship positions (in Ror)	98.380	98.544	98.538	99.368	98.791
	(4.906)	(5.261)	(5.600)	(5.124)	(5.279)
No. of students (in Ror)	23.700	22.711	24.156	25.730	24.095
	(14.204)	(14.354)	(13.991)	(14.091)	(14.223)
No. of high school graduates (in Ror)	26.081	25.755	27.289	27.758	26.775
	(6.313)	(6.526)	(6.218)	(7.064)	(6.673)
No. of Universities (in Ror)	10.789	9.620	10.916	11.381	10.585
	(10.304)	(9.666)	(10.037)	(9.988)	(9.938)
N	331(9.17%)	1'302(36.07%)	826(22.88%)	1'151(31.88%)	3'610

Notes: Cells contain sample means and standard deviations in parentheses. Intended investment levels in post-secondary education: no intention (0), apprenticeship (1), tertiary apprenticeship (2), or university (3). Individual characteristics are assessed at the time of answering the youth questionnaire (at age 17); p denotes the subjective completion belief; definitions of other variables are given in Appendix, Table B1.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

more risk-loving attitudes, but the differentials across categories are small.

Individual and family backgrounds are captured by individual's gender, number of siblings, whether they are second-generation immigrants (youths whose parents were both born in a foreign country), whether at least one parent has a college education, whether at least one parent is currently unemployed, and the logarithm of net household income. Intentions tend to be higher among males, children from smaller families, natives, youths with employed and college-educated parents, and youths with a higher household income.

The regional labor and education market (Ror) characteristics relevant for the students' choice sets are a mix of (exogenous) educational supply and demand shifters. We use the cyclical component of the youth unemployment rate, as well as the number of apprenticeship positions, students, high school graduates, and universities in the region. Intentions to invest in more academically ambitious tracks are related to a more favorable local labor and education market. However, youths with no investment intentions are not faced with clearly worse or better labor/education market conditions. In the following analysis, we will also account for region and year of first questioning (which is almost identical to students' age) fixed effects.[20]

3 Subjective completion beliefs and educational outcomes

3.1 Determinants of subjective completion beliefs

To analyze how the variables we discussed in the previous section relate to subjective completion beliefs, we estimate OLS regressions of the model

$$p_i = x_i'\beta^p + \upsilon_i, \tag{1}$$

where i indexes individuals, p_i is subjective completion belief, x_i are varying sets of explanatory variables with corresponding vector of coefficients β^p, and υ_i is an unobserved error term. The estimates are presented in Table 2.[21] In Column (1), the beliefs are explained solely by academic ability, year, and region fixed effects. In Column (2), we add the personality measures; in Column (3), individual and family background characteristics; and, finally, in Column (4), regional labor market measures.

The explained variation, as measured by the adjusted R^2, increases substantially only when academic ability and personality measures are included in the regressions, but is relatively unaffected when

[20] For some of the regressions, the number of students in a state is too small. To obtain consistent samples, we use a broader grouping by dividing Germany into the following five regions (and an indicator for missing values). *Southern Germany*: Baden-Wuerttemberg, Bavaria; *Eastern Germany*: Berlin, Brandenburg, Saxony, Saxony-Anhalt, Mecklenburg-Western Pomerania; *Central Germany*: Hesse, Thuringia; *Western Germany*: North Rhine-Westphalia, Rhineland-Palatinate, Saarland; *Northern Germany*: Bremen, Hamburg, Lower Saxony, Schleswig-Holstein. We present analogous results of our main Table 3 in Appendix, Table D5 were we use federal states fixed effects, as the jurisdiction over educational policies is at the federal state level. The results are very similar.

[21] Note that our dependent variable is a fraction. In the Appendix, Table D1, we present fractional response regressions (as in Papke and Wooldridge, 1996, 2008). The results are virtually indistinguishable from the OLS estimates.

Table 2: Determinants of subjective completion beliefs, OLS regressions

Dependent variable: p_i, subjective completion belief (mean=0.776, standard deviation=0.198)				
	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.028	0.028	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.029	0.027	0.029	0.029
	(0.015)	(0.015)	(0.015)	(0.015)
Rec: Intermediate Track (yes/no)	0.066	0.060	0.057	0.056
	(0.012)	(0.011)	(0.011)	(0.011)
Rec: High school (yes/no)	0.051	0.043	0.039	0.038
	(0.011)	(0.011)	(0.011)	(0.011)
In high school (yes/no)	0.004	0.002	-0.005	-0.005
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.024	0.023	0.023
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.006	0.005	0.005
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.005	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.006	0.007	0.007
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		-0.001	0.000	0.000
		(0.003)	(0.004)	(0.004)
Conscientiousness (std)		0.032	0.033	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.011	-0.011
			(0.007)	(0.007)
Nr. siblings			-0.003	-0.003
			(0.002)	(0.003)
Second-generation migrant (yes/no)			-0.009	-0.009
			(0.013)	(0.013)
Parent college-educated (yes/no)			0.009	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			0.002	0.002
			(0.013)	(0.013)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3'610	3'610	3'610	3'610
adj. R^2	0.057	0.117	0.121	0.120
Fixed effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
F (p-value)	28.912 (0.000)	18.354 (0.000)	16.482 (0.000)	16.498 (0.000)
Personality		✓	✓	✓
F (p-value)		26.835 (0.000)	26.509 (0.000)	26.616 (0.000)
Family Background			✓	✓
F (p-value)			2.458 (0.016)	2.380 (0.020)
Labor market				✓
F (p-value)				0.310 (0.907)

Notes: Cells contain coefficients from linear regressions of subjective beliefs on varying sets of covariates, in (1) only on academic and region and time fixed effects (coefficients not presented), (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics. Robust standard errors in parentheses. All regressions include indicator variables for missing values in any of the covariates, and region and year fixed effects. F (p-value) are test statistics and p-values of tests of joint significance of corresponding groups of variables.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

adding individual and family characteristics or regional youth labor market variables.[22] The joint significance tests for subsets of variables reported at the bottom panel of the table paint a similar picture: Academic ability and personality characteristics are highly significant across all regressions, and their associated F statistics in Column (4) are 16.5 and 26.6, respectively. Individual and family characteristics are jointly significant, but at the five percent level with an F statistic of 2.5. The local youth labor market characteristics are insignificant and their F statistic is just 0.3. Since these coefficient estimates are neither jointly nor individually significant, we omitted them from the table. Thus, Table 2 indicates that youths' subjective completion probabilities reflect mainly their past academic record and personality traits. Youths' socio-economic family backgrounds are only mildly related to their completion beliefs, and the state of the local youth labor and education market seems not to affect their beliefs at this stage.

Looking at the determinants individually, all academic ability variables are consistently positive and significant. Somewhat surprisingly, already being enrolled in high school does not alter students' subjective completion beliefs, which might be a result of conditioning on prior track recommendations. As hypothesized by Coleman and DeLeire (2003), the locus of control is a very important determinant of subjective completion beliefs throughout the regressions, both in magnitude and significance.[23] Risk attitudes do not matter once family characteristics are accounted for. Our regressions indicate that among the Big Five measures of personality, conscientiousness is the most influential in shaping subjective beliefs. This finding highlights the importance of conscientiousness for education outcomes (see, *inter alia*, Borghans, Meijers and Ter Weel, 2006). While we find little evidence that openness or neuroticism influence completion beliefs, the effect of extraversion is about half as large as conscientiousness, and the effect of agreeableness, in turn, is about half as large as extraversion. On average, females seem to have lower subjective completion beliefs. This estimate is, however, only marginally significant (at least conditional on personality and academic ability). Household income is positively and significantly related to subjective completion beliefs. The remaining estimated coefficients are insignificant and mostly very small in magnitude.

3.2 Education choice with subjective completion beliefs

In this section, we turn to our central question of how subjective completion beliefs measured at age 17 years relate to educational intentions, investments, and, finally, degree attainment. To fix ideas, let individual i's utility u_{ij} from choosing an uncertain post-secondary educational track ($j \geq 1$) be

$$u_{ij} = \begin{cases} \mu_{ij} + \nu_{ij} & \text{with probability } p_i \\ \bar{\mu}_{ij} + \nu_{ij} & \text{with probability } (1 - p_i), \end{cases} \tag{2}$$

[22] This results does not change when reversing the order in which the sets of variables are included.

[23] Related to this, Jaik and Wolter (2016) find that students with external locus of control have a higher intention to delay their educational transition. Caliendo, Cobb-Clark and Uhlendorff (2015) also find a strong link between subjective beliefs and the locus of control in the realm of job search among the unemployed. However, for an opposing view see Cebi (2007).

where p_i is subjective completion belief, μ_{ij} and $\bar{\mu}_{ij}$ are the utilities from completing and dropping out, respectively, and ν_{ij} is a utility component unaffected by completion.[24] The associated expected utility is

$$\begin{aligned} U_{ij} &= p_i\mu_{ij} + (1-p_i)\bar{\mu}_{ij} + \nu_{ij} \\ &= \bar{\mu}_{ij} + p_i(\mu_{ij} - \bar{\mu}_{ij}) + \nu_{ij}. \end{aligned} \tag{3}$$

Hence, adolescents get a baseline utility from attending a particular education track $\bar{\mu}_{ij}$. The subjective completion belief p_i weights the utility differential between completing and not completing up or down. Since not investing in education does not involve completion uncertainty, the utility is simply

$$U_{i0} = \mu_{i0} + \varepsilon_{i0}, \text{ with certainty.} \tag{4}$$

In this section, we assess the investment in any post-secondary education $U_{ij} = U_i$ for $j \geq 1$, against not investing U_{i0}, which directly relates to the survey question of p_i. A student prefers to invest in education if $U_i > U_{i0}$; where, by standard normalization, $\mu_{i0} = 0$. Taking averages across individuals, adding covariates that measure observed preferences and skills (i.e., $\nu_i = x_i'\beta^d + \varepsilon_i$),[25] and assuming that $\epsilon_i = \varepsilon_i - \varepsilon_{i0}$ follows a standard normal distribution, we estimate probit models of the form

$$d_i = 1[\alpha p_i + x_i'\beta^d + \epsilon_i > 0]. \tag{5}$$

We consider three different binary outcomes d_i: First, whether a student intends to complete any post-secondary education, which is measured concurrently with subjective beliefs at age 17 years. Then, the expectation of (5) gives $P(U_i > 0)$, so that $\alpha = \mu - \bar{\mu}$. A similar interpretation is possible when d_i represents the second outcome – a student actually invests in any post-secondary education, i.e., whether the student started an apprenticeship, a tertiary apprenticeship, or a university degree. This event can be a few months or a few years away. It then corresponds to the revealed preferences of actual investment in post-secondary education. Finally, the interpretation is somewhat different when d_i stands for the third outcome—the completion of an apprenticeship or a university degree, an event that is at least a couple of years away. In this case, α gives an indication of the student's ability to incorporate information beyond that in x_i into his or her forecast of $d_i = 1$, an interpretation of subjective beliefs along the lines of Finkelstein and McGarry (2006). The students process all their available information in forming their beliefs, meaning that relevant information over and above their subjective beliefs are either not used, not used efficiently, or influence the decision through a different channel from subjective completion uncertainty.

Table 3 contains the estimation results. Panels (A) to (C) present the probit regressions of subjective completion beliefs on different education outcomes. In each panel, we report the estimated coefficients, robust standard errors (in parentheses), average marginal effects (in squared brackets), pseudo R_n^2 for models estimated without and with p_i, and sample statistics for the respective subsamples. Columns

[24] This assumption implies that preferences are additively separable. This excludes interactions between uncertainty and other covariates, which is, however, a promising line of future research.

[25] We will discuss the role of unobserved preferences correlated with p_i in detail in the next section.

Table 3: Subjective completion beliefs and education outcomes

Dependent variables: Indicator variables for post-secondary education intentions (A), investment (B), or completion (C).							
	probit				probit eev		
					$\rho = \hat{\rho}^o$	$\rho = .1$	$\rho = .3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Intentions							
p	0.921	0.809	0.716	0.704	0.538	0.600	0.371
	(0.142)	(0.146)	(0.149)	(0.151)	(0.149)	(0.150)	(0.144)
	[0.140]	[0.121]	[0.107]	[0.103]			
R^2_n	0.029	0.040	0.048	0.057			
$R^2_n(p)$	0.049	0.055	0.059	0.067			
Sample: $N = 3'610$, $\bar{d} = 0.91$, $\bar{p} = 0.78$, $SD(p) = 0.20$, $\hat{\rho}^o = 0.157$							
(B) Actual investment							
p	0.997	0.915	0.902	0.862	0.842	0.758	0.523
	(0.223)	(0.228)	(0.239)	(0.249)	(0.249)	(0.248)	(0.238)
	[0.069]	[0.062]	[0.056]	[0.044]			
R^2_n	0.087	0.100	0.120	0.182			
$R^2_n(p)$	0.113	0.121	0.138	0.197			
Sample: $N = 2'116$, $\bar{d} = 0.96$, $\bar{p} = 0.77$, $SD(p) = 0.20$, $\hat{\rho}^o = 0.020$							
(C) Actual completion							
p	0.434	0.410	0.387	0.373	0.374	0.272	0.056
	(0.181)	(0.185)	(0.190)	(0.192)	(0.192)	(0.191)	(0.183)
	[0.172]	[0.162]	[0.153]	[0.148]			
R^2_n	0.089	0.093	0.102	0.121			
$R^2_n(p)$	0.092	0.096	0.104	0.123			
Sample: $N = 1'372$, $\bar{d} = 0.54$, $\bar{p} = 0.77$, $SD(p) = 0.20$, $\hat{\rho}^o = 0.000$							
Fixed effects	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓	✓	✓	✓
Personality			✓	✓	✓	✓	✓
Family Background				✓	✓	✓	✓
Labor market				✓	✓	✓	✓

Note: Table entries are coefficients (robust standard errors in parentheses; average marginal effects in squared brackets), from probit regressions (1)-(4) and probit regressions with endogenous explanatory variable (5)-(7). Variable p denotes the subjective completion beliefs. Other covariates: in column (1) in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (7) family background, individual, and local labor market variables. In probit eev regressions the correlation between errors, ρ, is restricted as indicated in column headers: (5) $\hat{\rho}^o$, the estimated selection-on-observables; (6) 0.10; (7) 0.30. R^2 and $R^2(p)$ are McFadden's pseudo-R^2 excluding and including p, respectively; $\bar{d}$ is the mean of the dependent variable, $\bar{p}$ and $SD(p)$ are mean and std. deviation of p_i.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

(1) to (4) contain simple probit estimates of education outcomes on subjective beliefs and varying sets of covariates: The specification in Column (1) contains, apart from p_i, only an indicator of whether the student is currently in high school, as well as region and year fixed effects. Thus, in this specification, any other variable acts on the intention to invest in education through its effect on p_i. The next Columns progressively control for sets of academic (Column 2), personality (Column 3), and family and labor market variables (Column 4). We turn to the results in Columns (5)-(7) in the next section.

Panel (A) contains results corresponding to the intention to invest in any post-secondary education. The uncertainty measured by p_i is an important predictor for investment intentions: The coefficients on subjective beliefs are large and highly significant throughout the probit regressions. The average marginal effects are economically relevant. In the most parsimonious specification, increasing subjective beliefs in the population by one standard deviation increases intentions to invest in post-secondary education by 2.8 percentage points (0.140×0.2), which is quite large relative to the nine percent of students who do not intend to invest in a post-secondary education. These figures change little if we condition on increasing sets of background characteristics commonly considered in the literature. A one standard deviation change results in an increase of 2.1 percentage points using all background characteristics. Moreover, the increase in the R^2 when including subjective beliefs in the regression in Column (1) (from 2.9 to 4.9 precent) is similar to the increase when adding the full sets of both academic and personality variables (from 2.9 to 4.8 precent). Thus, the predictive power of subjective beliefs is equal to the one of academic and personality variables combined. In sum, subjective beliefs are strongly related to intended behavior (a result consistent with, for example, Huntington-Klein, 2015*b*).

Several explanations can account for this contemporaneous correlation between beliefs and intentions. Therefore, we examine whether the link from beliefs to intention carries over to revealed preferences in actual investments at least two years later. In Panel (B), our dependent variable is now an indicator that equals one if the student started any post-secondary education. Compared to the sample used in (A), the sample in (B) only excludes students who are still in school and students who have not completed any subsequent questionnaires two years after the youth questionnaire at age 17. The average marginal effects are somewhat smaller in absolute value than for the intentions, ranging from 1.4 to 0.9 percentage points for a one standard deviation increase in subjective beliefs. The effects are of similar magnitude as before when considered relative to the group who fails to invest in post-secondary education, which comprises about four percent of youths: A one standard deviation increase in p_i is associated with a predicted reduction of this group by between 34 percent to 22 percent.[26] This result suggests that subjective completion beliefs not only drive hypothetical, intended investment, but have real behavioral consequences. Compared to the previous results on students' intentions, the set of family background and labor market variables explain a somewhat larger fraction of the completion belief effect and exhibit more explanatory power.

Finally, we compare how subjective completion beliefs relate to actual completion at least five years later. This correlation can be interpreted as the degree to which students can predict their future

[26] In Table D2, we show that similar results are obtained when restricting the sample to students who expressed earlier intentions in post-secondary education.

outcomes.[27] The estimation results are given in Panels (C). The average completion rate is roughly 54 percent. Unconditionally, a one standard deviation increase in subjective beliefs increases completion rates in the population by 3.4 percentage points. This effect shrinks to 3.0 percentage points when including the full set of individual, family, and regional characteristics. Thus, a one standard deviation increase in p_i is predicted to reduce the group of students dropping out of post-secondary education by about 6.5 percent. For completing a post-secondary education, the explanatory power of family background and labor market characteristics is substantial relative to that of the other covariates. For instance, relative to a baseline of only fixed effects, the inclusion of p_i, academic, and personality variables adds successively about three, three, and eight percent to the R^2, while family and labor market variables add another 18 percent. This result stands in contrast to the results from the previous section, and suggests that students do not optimally take thier family and local labor market information into account when forming their beliefs.

3.3 Robustness and Altonji-Elder-Taber bounds

Taken together, the results show that subjective completion beliefs formed during secondary education are predictive over a long time horizon, which is consistent with substantial inertia due to preexisting beliefs in these choices, as was hypothesized in the studies cited above. Moreover, these early subjective completion beliefs are predictive even after accounting for a comprehensive set of previously identified, important characteristics. In the appendix, we present further results showing the robustness of these findings across a number of alternative specifications.

First, we show that the results change little when using only students who expressed positive intentions, and additionally control for the subjective investment probability (cf. Table D2). We are therefore confident that the measure corresponds well to the conditional subjective probability of completing an educational degree. Next, we show that the results are robust to dichotomizing the subjective belief to a dummy variable (cf. Table D3), thus accounting for potential non-linearity, as discussed in Pinger (2015).

A key determinant is GPA, which the literature has found to be the main driver of learning about own ability among college students. To show that our results do not hinge on the measure of GPA we use, we assess various alternative standardizations and specifications. First, we standardize GPA within high schools, since grades might be based on different standards across school tracks (Table D5), and within federal states, which accounts for potentially different grading standards across state education systems (Table D6). Further, we use a fifth-order polynomial in GPA to demonstrate that the effects of beliefs are not spuriously picking up non-linearities in academic ability (Table D6). Additionally, we use federal state dummies (Table D6) instead of the region dummies used in the main specification

[27] Since completing a program and graduating takes some time, we consider only students whom we see at least five years after they have taken the youth questionnaire. Students who were interviewed in earlier years are more likely to have completed their degrees simply by virtue of having been in the sample for a longer period of time. However, this mechanism is captured by the year fixed effects, and is therefore unlikely to bias our results. A second concern is that some of the observations are censored: As of the time we observe them, some students have not yet completed their degree, but they might do so in the future. In this sense, our results should be interpreted as representing the average effect of completion beliefs on completion within a given time frame.

(cf. footnote 20). Finally, we present separate estimations for students enrolled in high school when answering the youth questionnaire to account in a completely flexible way for the different default choices discussed above (Table D4).

Although the results in Columns (1)-(4) in Table 3 show that the coefficient on p_i does not change much even after including a very large set of potential confounders, a remaining concern might be that the effect of subjective beliefs, α, is confounded with further unobserved variables, i.e., in terms of our empirical model, that there are unobserved variables in ϵ_i that are correlated with p_i. In order to show that our results are robust to this selection on unobservables, we go one step further and use a bounding strategy introduced by Altonji, Elder and Taber (2005*a*,*b*, 2008, hereafter, AET) for the coefficients by taking into account potential selection on unobservable tastes and preferences for education when estimating (5). We show how α varies if there is a correlation ρ between unobservable components of d_i and p_i, and we present a suggestive upper bound under the assumption that the correlation between observables is informative about the correlation between unobservables. In contrast to AET, our main variable is a fraction rather than an indicator. Instead of estimating a bivariate probit, we therefore estimate a probit with continuous endogenous explanatory variable, or "probit eev".[28] Note that Column (4) of Table 2 in combination with Column (4) of Table 3 is equivalent to the probit eev with $\rho = 0$; hence, the probit eev model nests the two separate models above. More detailed information on the probit eev estimation and the AET bound is provided in Appendix A.1. Column (5) constrains ρ to be equal to the selection-on-observables $\hat{\rho}^o$. Finally, Columns (6) to (7) contain the probit eev estimates using the full set of covariates and ρ constrained to 0.1 and 0.3, respectively. All coefficients are positive up to a correlation of 0.3 (which is more than double the implied upper bound based on selection on observables), and the coefficients also remain statistically significant when the outcomes of interest are intentions and investments. This is a sizeable correlation when comparing it to the applications considered in AET, and to the selection on observables estimated in our data. Thus, the results appear to be robust to substantial selection on unobservables.

4 A sequential model of educational choice with completion beliefs

In this section, we develop and estimate a model of sequential human capital investment along the lines of Taber's (2000; 2001) seminal contribution. Our model is deliberately held simple, but it is capable of accounting for various features shown to be relevant in the literature. First, we allow for the sequential nature of the process: students can only decide whether they want to go to university if they previously chose to finish high school (Altonji, 1993; Comay, Melnik and Pollatschek, 1973). Second, we introduce the dynamics of the optimization process: when deciding whether to enter the labor market or to continue with high school, forward-looking students account for the option value of continuing education after finishing high school (Stange, 2012; Trachter, 2015). Finally, we allow

[28] The use of a continuous normal variable is motivated by the estimation of (1), where we found that it made little difference whether the equation was estimated by OLS or a fractional response model (cf. Appendix, Table D1). As a robustness check, we dichotomize the subjective beliefs at $p \geq 70$ percent and estimate bivariate probit regressions as in AET. Estimates for such an approach can be found in the Appendix, Table D3. The results are qualitatively similar but somewhat attenuated, which is likely due to the artificial measurement error introduced by the dichotomization.

for unobserved factors that influence students' choice utilities, which may be correlated across choices and over time, and which is a topic of substantial attention in the returns to education literature (see, e.g., Belzil, 2007; Card, 2001).

The implementation of such a model is not without problems, arising from limitations of the survey data we use. Ideally, a repeated measure of subjective beliefs, elicited at the end of high school, would make it possible to study students' learning about their own ability during high school. Lacking repeated measures, we focus here on the relevance of *ex ante* completion beliefs. Second, hypothetical beliefs for completing each education track are not elicited in the survey, hampering our ability to study track-specific uncertainty. Keeping this in mind, we interpret the beliefs as a measure of overall post-secondary completion uncertainty.

4.1 Empirical Model

We consider a stylized two-period model in which students sequentially choose between risky education paths, as outlined in Figure 3. *Ex ante*, students do not know for certain whether they will successfully complete the chosen education track, but they have subjective beliefs, p_i, about finishing. The first period or first stage ($T = 1$) occurs when students finish compulsory education at the age of 17 years. At this point, they face the choice between dropping out of school ($d_{i1} = 0$), investing in an apprenticeship training ($d_{i1} = 1$), or continuing with high school education ($d_{i1} = 2$). A high school degree involves the option value of continuing with tertiary education. Students who choose high school reach the second stage ($T = 2$), after which they have the choice of either investing in a tertiary apprenticeship ($d_{i1} = 2, d_{i2} = 0$) or in a university education ($d_{i1} = 2, d_{i2} = 1$).[29]

As mentioned above, apprenticeships, tertiary apprenticeships, and university all involve uncertainty, which we model according to equations (2)-(4). By backward induction, we begin with the students' second stage problem. Students advancing to the second stage choose between starting a tertiary apprenticeship ($j = 2$) or going to university ($j = 3$). We denote this choice by d_{i2}, a binary variable where 1 represents choosing university,

$$d_{i2} = \begin{cases} 1 & \text{if } U_{i3} - U_{i2} > 0 \\ 0 & \text{if } U_{i3} - U_{i2} \leq 0, \end{cases}$$

which we specify, analogously to equation (4), as

$$\begin{aligned} U_{i2} &= \alpha_2 p_i + x'_{i,t+1}\beta_2 + \delta_2\theta_i + \epsilon_{i2} \equiv z_{i2,t+1} + \epsilon_{i2}, \\ U_{i3} &= \alpha_3 p_i + x'_{i,t+1}\beta_3 + \delta_3\theta_i + \epsilon_{i3} \equiv z_{i3,t+1} + \epsilon_{i3}. \end{aligned}$$

Here, $x_{i,t+1}$ consists of the same covariates considered above. Covariates that vary over time, such as some family characteristics (household income and parental unemployment status) and all labor market conditions, are measured two years after the youth questionnaire was answered, which is the

[29] In principle, students could also drop out at this point, but this is an extremely rare event in the data and therefore not modeled (see also Fossen and Glocker, 2014).

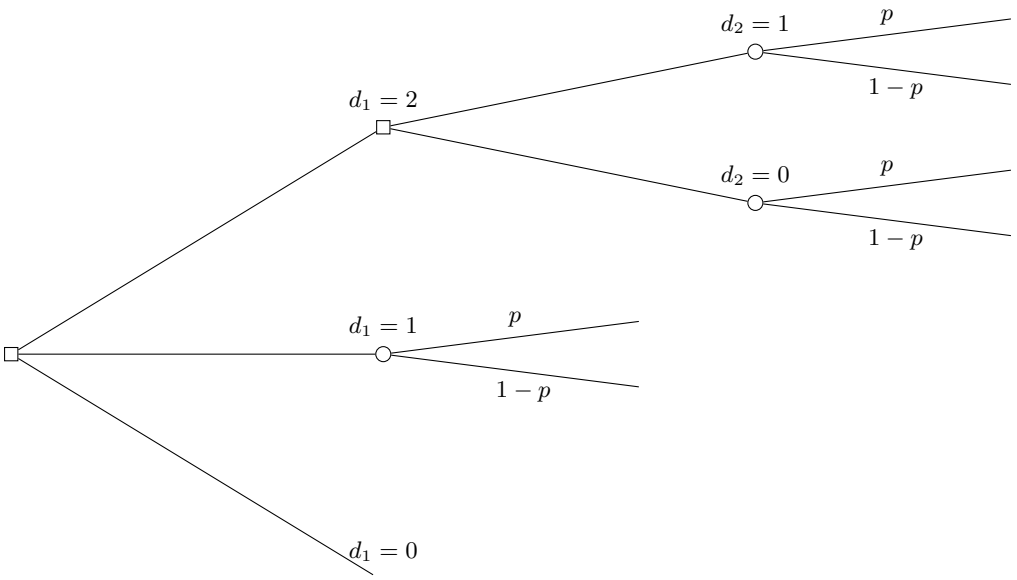


Figure 3: Sequential education decisions and timing of events

time one would need to start tertiary education after obtaining a high school degree.[30] This exogenous variation in the decision problem induced by the timing of events provides an additional source of identification (see French and Taber, 2011; Taber, 2000, for a discussion on the identification for these models), which has become standard practice in the literature on dynamic models of education choice (e.g., Heckman et al., 2014; Taber, 2001). To allow for dependence of the unobservables between the two time periods in a flexible way, we add a standard normal random variable θ_i to the utilities, capturing unobserved tastes and preferences for education. We assume that $\epsilon_{i3} - \epsilon_{i2} \sim N(0, \sigma^2)$; thus the probability of choosing university is given by

$$P(d_{i2} = 1) \quad = \quad \Phi\left(\frac{z_{i3,t+1} - z_{i2,t+1}}{\sigma}\right),$$

where $\Phi(\cdot)$ represents the univariate normal cdf.

In the first stage, the student has an expectation about her second stage decision (she knows the distribution of $\epsilon_{i3} - \epsilon_{i2}$) but does not know her realized value. If students knew their realized $\epsilon_{i3} - \epsilon_{i2}$ at the time of the first stage, the model would reduce to a static polychotomous choice problem.[31] Thus, a student's expectation about her value of advancing to the second stage, as formed during the first stage, is

$$\mathrm{E}(\max(U_{i3} - U_{i2}, 0)) \quad = \quad z_{i2,t} + \sigma\left[\Phi\left(\frac{z_{i3,t} - z_{i2,t}}{\sigma}\right)\frac{z_{i3,t} - z_{i2,t}}{\sigma} + \phi\left(\frac{z_{i3,t} - z_{i2,t}}{\sigma}\right)\right] \equiv EV_{i,t},$$

[30] We use students' location at the age of 17 for the region fixed effects and the regional characteristics, to avoid a bias due to moving.

[31] We show a simple multinominal probit model for intentions in the appendix, Table C2. Table C3 additionally presents simple probit regressions of investment in the different educational tracks conditional on intentions. The results, where comparable, are in line with those of the structural model.

and $\phi(\cdot)$ denotes the normal pdf.[32] Corresponding to the student's information set and time-varying family background, labor market and education supply and demand characteristics are now measured at time t, one year before the student answers the youth questionnaire. The difference between utility from continuing with high school and utility from dropping out of high school is then

$$\begin{aligned} U_{iHS} - U_{i0} &= \alpha_{HS} p_i + x'_{i,t}\beta_{HS} + \delta_{HS}\theta_i + EV_{i,t} + \epsilon_{iHS} \\ &\equiv z_{iHS,t} + EV_{i,t} + \epsilon_{iHS}, \end{aligned}$$

which comprises EV_i, the option value of continuing to the second stage. Finally, the apprenticeship utility is

$$U_{i1} - U_{i0} = \alpha_1 p_i + x'_{i,t}\beta_1 + \delta_1\theta_i + \epsilon_{i1} \equiv z_{i1,t} + \epsilon_{i1}.$$

From the two preceding equations, it is clear that first stage choices are made relative to the baseline utility U_{i0}, which corresponds to dropping out of high school (i.e., not investing in post-secondary education). The model requires two additional normalizations. First, since the unobserved heterogeneity (θ_i) has no natural scale, we set $\delta_1 = 1$. Thus, the impacts of unobserved heterogeneity on second stage utilities, δ_2 and δ_3, are then estimated relative to a one standard deviation impact on high school utility. Second, we normalize $z_{iHS,t} = 0$ since z_{iHS}, z_{i2}, and z_{i3} are not separately identified.[33] This means, for example, that α_2 estimates the combined impact of uncertainty on high school and tertiary apprenticeship utility; and α_3, the combined impact on high school and university utility.

Then, since the ϵ's are normally distributed, the first stage choice probabilities are

$$\begin{aligned} P(d_{i1} = 2) &= \Phi_2\left(EV_{i,t},\ \ EV_{i,t} - z_{i1,t}\right), \\ P(d_{i1} = 1) &= \Phi_2\left(z_{i1,t},\ \ z_{i1,t} - EV_{i,t}\right), \\ P(d_{i1} = 0) &= 1 - P(d_{i1} = 1) - P(d_{i1} = 2), \end{aligned}$$

where $\Phi_2(\cdot)$ stands for the bivariate standard normal cdf. The individual likelihood contribution, conditional on the unobserved heterogeneity θ_i, is given by

$$\begin{aligned} l_i(\theta_i) = &\left\{1 - P(d_{i1} = 1) - P(d_{i1} = 2)\right\}^{\mathbf{1}(d_{i1}=0)} \times \left\{P(d_{i1} = 1)\right\}^{\mathbf{1}(d_{i1}=1)} \\ &\times \left\{P(d_{i1} = 2)[1 - P(d_{i2} = 1)]\right\}^{\mathbf{1}(d_{i1}=2, d_{i2}=0)} \\ &\times \left\{P(d_{i1} = 2)P(d_{i2} = 1)\right\}^{\mathbf{1}(d_{i1}=2, d_{i2}=1)}, \end{aligned} \tag{6}$$

and to obtain the marginal likelihood contribution, we integrate over the distribution of θ,

$$l_i = \int l_i(\theta_i)\phi(\theta_i)d\theta_i,$$

[32] Further information on the model and the derivation of the EV_i are discussed more explicitly in Appendix A.2.

[33] Specifically, without further assumptions or restrictions, we cannot distinguish, for instance, between the first, additive part of EV_i (i.e., $z_{i2,t}$), and the first, additive part of $U_{iHS} - U_{i0}$ (i.e., $z_{iHS,t}$).

an expression which we approximate by simulation, $\tilde{l}_i$, by taking 100 random draws from the distribution of θ_i. We then maximize the simulated sample log-likelihood $\sum_i \ln \tilde{l}_i$.[34]

4.2 Estimation Results

The estimation results are depicted in Table 4 in two panels. The panel on the left contains estimates from a constrained version of the model without heterogeneity ($\theta_i = 0$ for all i), whereas the panel on the right contains estimates from the full model with unobserved heterogeneity. Moving from left to right, the Columns again contain the expanding set of covariates considered previously. With the exception of the family and regional labor market determinants, all the regressors are time-invariant.[35]

Table 4 depicts large and significant estimates for the coefficients of the subjective probabilities for both $d_{i1}=1$ and the two panels with $d_{i1}=2$. Thus, these results, too, suggest that a higher p_i pushes students away from leaving compulsory education without further human capital investments. In particular, the coefficients for $d_{i1}=2$ suggest that subjective completion beliefs are important determinants of second-stage participation – of completing high school and beginning a tertiary apprenticeship or university study (which confirms and extends the results found by Pinger, 2015, who considered high school enrollment). On the other hand, the coefficients for university and tertiary apprenticeship converge to similar estimates as the set of included covariates gets larger. A striking result is the drop in the subjective belief coefficient when controlling for academic ability among youths choosing university: from 1.77 to 1.25 (Columns 1 and 2) or from 2.01 to 1.47 (Columns 5 and 6). This decrease in the coefficients suggests that a large part of the belief effect is driven by academic ability.[36] Such a dramatic change in the estimates is not observed in the apprenticeship categories. A potential explanation for this is that original beliefs change in response to new information revealed by high school grades, which would be consistent with substantial belief updating, as documented by Stinebrickner and Stinebrickner (2014*a*) and Wiswall and Zafar (2015*a*) for students enrolled in college.

We now turn to the role of unobserved preferences for post-secondary education or unobserved skills, θ_i. Comparing the two panels of Table 4, we see that all the significant coefficients in the panel on the right, which accounts for such heterogeneity, are somewhat larger than the ones from the panel on the left. Recall that θ_i has no natural scale—its scale has been fixed such that a unit coefficient in the index corresponds to apprenticeship. The estimated coefficients for θ_i in the tertiary apprenticeship and university categories are highly significant and quite similar in magnitude, between 0.84 and 0.85 across all specifications (5)-(8). The results show that there is a strong positive correlation between unobserved preferences for any post-secondary education. Unobserved preferences for education are very important for adolescents' investment decisions, as found in prior literature (e.g. Bulman, 2015; D'Haultfoeuille and Maurel, 2013; Huntington-Klein, 2015*a*; Wiswall and Zafar, 2015*a*). On the other

[34] We used antithetic random draws as well as larger numbers of draws with little impacts on the estimates.

[35] The parameter σ is only identified when time-varying covariates are included (French and Taber, 2011; Taber, 2000). It is estimated only in Columns (4) and (8), while it is constrained to 1 otherwise. To further facilitate identification, the number of universities and the number of university students are only included in the second stage and EV. Results are robust to the inclusion of these variables in the first stage.

[36] This drop is also observed in the reduced form regression in the Appendix, Table C3.

Table 4: Dynamic models of actual investment

Dependent variables: Jointly estimated sequential model of investment in apprenticeship, tertiary apprenticeship, and university. Base category: no investment.								
	Dynamic model				Dyn. model with unobs. heterogeneity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Apprenticeship ($d_1 = 1$)								
p	1.173	1.213	1.193	1.151	1.406	1.432	1.415	1.354
	(0.303)	(0.314)	(0.329)	(0.341)	(0.313)	(0.324)	(0.340)	(0.352)
$\theta = 1$								
Tertiary apprenticeship ($d_1 = 2, d_2 = 0$)								
p	1.307	1.143	1.088	1.078	1.549	1.365	1.312	1.282
	(0.388)	(0.406)	(0.427)	(0.393)	(0.396)	(0.414)	(0.436)	(0.403)
θ					0.846	0.846	0.849	0.853
					(0.061)	(0.062)	(0.062)	(0.052)
University ($d_1 = 2, d_2 = 1$)								
p	1.769	1.248	1.163	1.046	2.012	1.468	1.385	1.249
	(0.348)	(0.367)	(0.388)	(0.488)	(0.357)	(0.376)	(0.397)	(0.495)
θ					0.840	0.840	0.842	0.844
					(0.048)	(0.051)	(0.052)	(0.068)
σ				0.208				0.224
				(0.184)				(0.187)
N	2'116	2'116	2'116	2'116	2'116	2'116	2'116	2'116
Fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓		✓	✓	✓
Personality			✓	✓			✓	✓
Family+Labor market				✓				✓

Note: Table presents estimates of the model in equation (6). In the panel "Dynamic model", $\theta_i = 0$ for all i. The model in the panel "Dyn. model with unobs. heterogeneity" estimated by MSL with 100 random draws from $N(0, 1)$. The sets of covariates correspond to those in Table 3. Standard errors in parentheses. All regressions include an indicator for being in high school with 17, region and time fixed effects.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

hand, there is no evidence for differences between unobserved tertiary apprenticeship-specific skills versus university-specific skills, again possibly a result of preference updating in high school.

4.3 Effect heterogeneity in academic ability and unobserved heterogeneity

Is there heterogeneity in the relevance of subjective uncertainty for different students? Figure 4 uses the results from Table 4 to visualize how the effect of p_i might differ for different "types" of students. Based on the estimates of the full specification from Column (8), we define four types by their academic ability level (high versus low GPA) and their unobserved skill level (high versus low θ_i). Low and high levels of GPA and θ_i are defined as the 25th and 75th percentiles in these variables' distributions. Their predicted educational investment choice probabilities (including the probability of no investment) are plotted against p_i, evaluated at sample means ($\bar{x}_i'\beta_j$). For students who have both high observed and unobserved skills (lower right panel), subjective completion beliefs have negligible effects on investment probabilities. Yet, for adolescents with low unobserved skills and high GPA (lower left), subjective beliefs positively influence all educational tracks. For students with low academic performance, subjective beliefs are more relevant if they also have a low preference for education (top left). In general, high GPA (for any level of θ_i) has a much larger effect on investment than high unobserved skills and preferences (for any level of GPA). Figure C1 contains a similar graph to Figure 4, for the marginal students (evaluated at $x_i'\beta_j = 0$), thus yielding even more pronounced effects. In sum, these findings suggests that subjective beliefs are most relevant for students with low academic ability and low unobserved skills, a group largely ignored in the current literature.

4.4 Subjective Beliefs, and direct vs indirect effects on investment

The final question we address is the extent to which changes in covariates are mediated by subjective beliefs. We focus on youths' investment in any post-secondary education because adolescents at risk of dropping out of school after finishing compulsory education are a key policy target population and, based on the results above, this is the margin for which subjective completion beliefs seem to matter most.[37] To disentangle the effects, we calculate average changes in the predicted investment probability for (i) a *ceteris paribus* change in a covariate of interest, and (ii) its corresponding *mutatis mutandis* change. The latter is the sum of the covariate's direct *ceteris paribus* change plus the indirect change that the covariate has through p_i. The results for selected variables are presented in Figure 5, using the estimates from Table 4, Column (8) (results for all variables can be found in Appendix, Table C1). We present average changes for two populations: all students, and a "low ability population".[38] While the predicted investment distribution of all students in Panel (A) is virtually identical to the actual distribution in our sample and to the official statistics in Germany, the low ability students in Panel (B) are at much higher risk of not investing in post-secondary education; if they invest, their

[37] Additionally, the results we present here are directly comparable to our reduced form estimates from Section 3.

[38] For "all students" we average over the entire estimation sample and use 100 draws for each student from θ_i with mean 0. For the "low ability population" we average only over the subsample of students with $GPA \leq -0.5$ and use 100 draws from θ_i with mean -1.

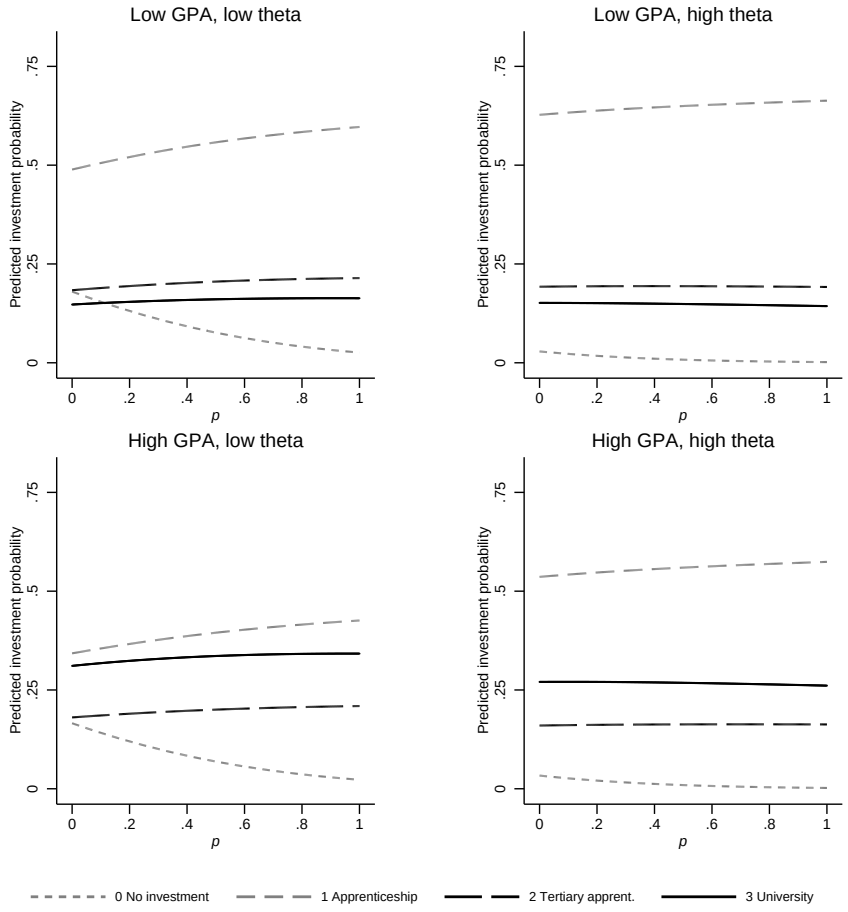
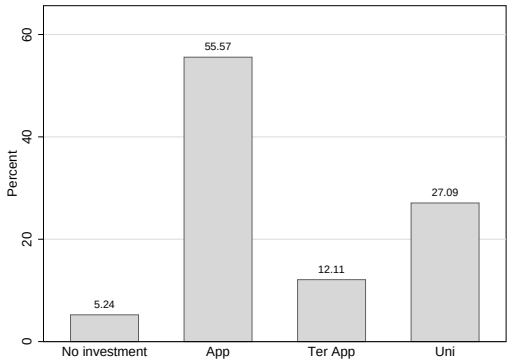


Figure 4: The role of academic ability and unobserved heterogeneity
Notes: Predicted probabilities constructed using estimates from Column (8) of Table 4, evaluated at sample means ($\bar{x}_i'\beta_j$). High GPA and θ are evaluated at the 75th percentiles; low GPA and θ, at the 25th percentile.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

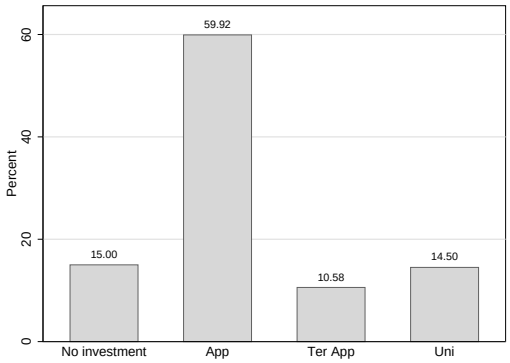
chosen path is much more likely to be vocational.

The two panels at the bottom of Figure 5, Panels (C) and (D), show the relative reduction in the predicted probability of not investing in any post-secondary education, for the whole population and the low ability population, respectively. We compare three covariates that have been highlighted in the literature as key determinants of human capital accumulation: GPA, a measure of cognitive ability; the (internal) locus of control, a measure of non-cognitive ability; and an indicator for having at least one college-educated parent, a measure of socio-economic status. As a reference, the panels also depict the direct effect of the subjective completion probability, p_i. We consider changes of one standard deviation in each of these variables, with the exception of having at least one college-educated parent, for which we consider a unit change. The darker bars denote the direct (or *ceteris paribus*) effects, while the lighter bars depict the portion of the total effect that works indirectly through increasing the completion beliefs. Hence their portion indicates the "importance" of the indirect channel through p_i.

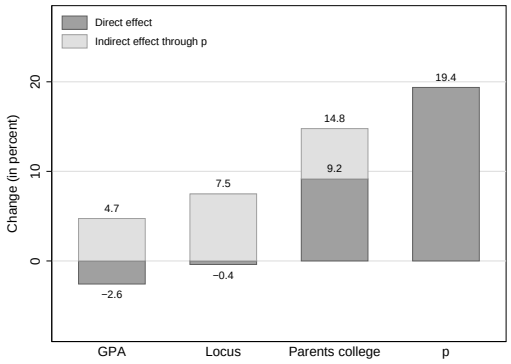
All variables shown in the panels reduce the probability of not investing in any education. Our model implies that an increase in students' GPA by one standard deviation would increase their dropout risk by 2.6 percent, assuming this change left the students' subjective beliefs unaffected. However, an increased GPA comes hand in hand with an increase in subjective belief that in turn reduces the



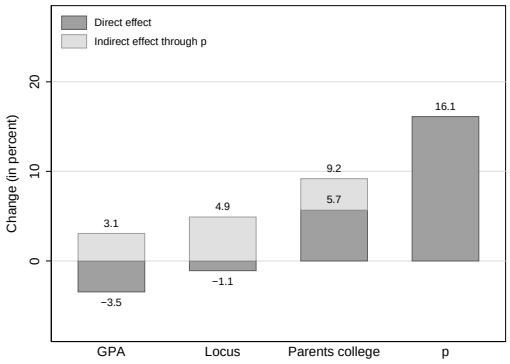
(A) Predicted investment distribution: all students



(B) Predicted investment distribution: low ability students



(C) Predicted % reduction in no investment: all students



(D) Predicted % reduction in no investment: low ability students

Figure 5: Comparison of interventions

Notes: All predictions constructed using estimates from Column (8) of Table 4. Figures (A) and (C), "all students": averages over entire sample, θ_i integrated out with mean equal 0. Figures (B) and (D), "low ability students": averages over individuals with GPA$\leq$ 0.5, with mean unobserved skills equal to -1. Top Figures (A), (B): average predicted probabilities for investment in post-secondary education. Bottom Figures (C), (D): predicted reductions, in %, in the probability not to invest for a one standard deviation increase in GPA, locus of control, p, and for a unit increase in parents college. Numbers on top of the bars denote the total change, those below the indirect effect, thus their difference is the indirect effect.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

dropout risk by 7.3 percent, leaving a total reduction in dropout risk of 4.7 percent. The fact that GPA has no substantial direct effect on dropout risk does not imply that GPA is unimportant for education investments. Figure C2 in the Appendix documents how changes in GPA and other variables are indeed associated with substantial direct shifts in the distribution of predicted investment. However, these shifts all take place at the intensive margin, that is, between different types of post-secondary education (and for GPA, in particular, a major shift from apprenticeships towards university). A similar picture emerges for the locus of control, which has only negligible direct effects on investment when beliefs are held constant. Again, its total effect is large and positive, reducing dropout by 7.5 percent in the total student population and by 4.9 percent among the low ability and low preference population. Hence, this effect is driven entirely by its impact on subjective beliefs. We interpret this result as strong evidence for the hypothesis of Coleman and DeLeire (2003, p3), that "locus of control operated through teenagers' expectations of the returns to human capital investments". Finally, roughly one third of the intergenerational effect, captured by having at least one college-educated parent, is estimated to be due to the increasing subjective completion belief. Taken together, these results suggest that policies aimed at increasing take-up of post-secondary education for at-risk youths

might be most effective if they target, or at least take into account, students' completion beliefs.

5 Conclusion

In this paper, we investigated the role of uncertainty about the likelihood of completion in youths' post-secondary education choices using their elicited subjective beliefs about successfully finishing their chosen post-secondary education. The students' young age and the long time horizon make this an especially hard problem, and it is remarkable that these necessarily crude initial beliefs retain their predictive power over several years. The effects of subjective beliefs on investment intentions and actual investments in post-secondary education are substantial, and remain so even after controlling for a large set of observable characteristics, and bounding against unobservables. Moreover, subjective beliefs have an explanatory power comparable to that of academic and personality variables combined. Finally, our results indicate that subjective probabilities of completion are also predictive of actual completion. Thus, students' beliefs contain private information currently not captured in empirical human capital models. Conversely, our results also suggest that students disregard some relevant information when forming their beliefs, such as family and education market characteristics.

The results from the dynamic sequential model with unobserved heterogeneity shed light on some additional aspects of educational choice. For one, the sequentiality of choices shows that subjective completion beliefs have a highly significant effect on the combined high school and second-stage choices. Furthermore, we have seen that accounting for the option value and for unobserved skills can modify the salience of the effect of beliefs on the choice probabilities. Differentiating between post-secondary education choices, we find that subjective beliefs are most relevant for students who aim for a university degree. This is most likely due to the information revealed by GPA, which broadly confirms results found in the literature. Most notably, we confirm Zafar's (2011*b*) finding that *ex ante* subjective beliefs continue to be important until the degree is completed. Advancing his findings, we conclude that this is true for subjective beliefs formed in or even before investing (or staying) in high school.

After conditioning on academic ability, subjective beliefs are most relevant for students who start an apprenticeship. Our results suggest that subjective beliefs are of crucial importance for students with lower academic ability and low unobserved preferences or skills for post-secondary education, who primarily invest in apprenticeships. These students have been largely ignored in the present literature on subjective beliefs in education choice, and evidence on their learning/decision-making processes is scant. Our study suggests that these students deserve much more attention, especially since apprenticeship systems are now tested or implemented in several countries (e.g., President Obama's State of the Union Address, 2014).

References

Almlund, Mathilde, Angela Lee Duckworth, James Heckman and Tim Kautz. 2011. Personality Psychology and Economics. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 1–181.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11(1):48–83.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*a*. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources* 40(4):791–821.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005*b*. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1):151–184.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2008. "Unobservables Assess Bias from Using Selection on Observed Variables to when Evaluating Swan-Ganz." *American Economic Review: Papers and Proceedings* 98(2):345–350.

Arcidiacono, Peter, Joseph V. Hotz, Arnaud Maurel and Teresa Romano. 2014. "Recovering Ex Ante Returns and Preferences for Occupations using Subjective Expectations Data." *National Bureau of Economic Research. Working Paper Series Nr. 20626*.

Arcidiacono, Peter, Joseph V. Hotz and Songman Kang. 2013. "Modeling college major choices using elicited measures of expectations and counterfactuals." *Journal of Econometrics* 166(1):3–16.

Attanasio, Orazio P. and Katja M. Kaufmann. 2014. "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics* 109:203–216.

BBSR, Bundesinstitut für Bau-Stadt-und Raumforschung. 2013. "Raumordnungsregionen (ROR)." `http://www.bbsr.bund.de/BBSR/DE/Raumbeobachtung/Raumabgrenzungen/Raumordnungsregionen/raumordnungsregionen_node.html`

Belzil, Christian. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51(5):1059–1105.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: The H&R Block FAFSA Experiment." *Quarterly Journal of Economics* 127(3):1205–1242.

Bond, Timothy N., George Bulman, Xiaoxiao Li and Jonathan Smith. 2016. "Updating Human Capital Decisions: Evidence from SAT Score Shocks and College Applications." *MPRA Paper No. 72497*.

Borghans, Lex, Angela L. Duckworth, James J. Heckman and Bas ter Weel. 2006. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43(4):972–1059.

Borghans, Lex, Huub Meijers and Bas Ter Weel. 2006. "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46(1):2–12.

Bound, John and Sarah E. Turner. 2011. Dropouts and Diplomas. The Divergence in Collegiate Outcomes. In *Handbook of the Economics of Education*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann. Vol. 4 Elsevier B.V. pp. 573–613.

Bulman, George. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics* 7(4):1–36.

Caliendo, Marco, Deborah A. Cobb-Clark and Arne Uhlendorff. 2015. "Locus of Control and Job Search Strategies." *Review of Economics and Statistics* 97(1):88–103.

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5):1127–1160.

Carneiro, Pedro, Karsten T. Hansen, James J. Heckman and Pedro Carneiro. 2003. "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice." *International Economic Review* 44(2):361–422.

Castex, Gonzalo. 2015. "College risk and return." *Working Paper, Central Bank of Chile*.

Cebi, Merve. 2007. "Locus of Control and Human Capital Investment Revisited." *Journal of Human Resources* 42(4):919–932.

Christofides, Louis N., Michael Hoy, Joniada Milla and Thanasis Stengos. 2015. "Grades, aspirations and post-secondary education outcomes." *Canadian Journal of Higher Education* 45(1):48–82.

Cobb-Clark, Deborah A. 2015. "Locus of control and the labor market." *IZA Journal of Labor Economics* 4(3):1–19.

Coleman, Margo and Thomas DeLeire. 2003. "An Economic Model of Locus of Control and the Human Capital Investment Decision." *Journal of Human Resources* 38(3):701–721.

Comay, Yochanan, Arie Melnik and Mosche A. Pollatschek. 1973. "The Option Value of Education and the Optimal Path for Investment in Human Capital." *International Economic Review* 14(2):421–435.

Costa, Paul Jr. and Robert McCrae. 1992. "Four ways five factors are basic." *Personality and Individual Differences* 13(6):653–665.

Cunha, Flavio and James J. Heckman. 2007. "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling." *Labour Economics* 14(6):870–893.

Cunha, Flavio, James J. Heckman and Salvador Navarro. 2005. "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers* 57(2):191–261.

D'Haultfoeuille, Xavier and Arnaud Maurel. 2013. "Inference on an extended Roy model, with an application to schooling decisions in France." *Journal of Econometrics* 174(2):95–106.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde. 2010. "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review* 100(3):1238–1260.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner. 2011. "Individual risk attitudes: Measurement, determinants, and behavioral consequences." *Journal of the European Economic Association* 9(3):522–550.

Finkelstein, Amy and Kathleen McGarry. 2006. "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market." *American Economic Review* 96(4):938–958.

Foley, Kelly, Giovanni Gallipoli and David A. Green. 2014. "Ability , Parental Valuation of Education and the High School Dropout Decision." *Journal of Human Resources* 49(4):906–944.

Fossen, Frank M. and Daniela Glocker. 2014. "Stated and revealed heterogeneous risk preferences in educational choice." *IZA DP 7950*.

French, Eric B. and Christopher R. Taber. 2011. "Identification of models of the labor market." *Handbook of Labor Economics* 4:537–617.

Greene, William H. 2012. *Econometric Analysis*. 7 ed. Pearson Education Limited. Edinburgh Gate, Harlow.

Hahn, Elisabeth, Juliana Gottschling and Frank M. Spinath. 2012. "Short measurements of personality - Validity and reliability of the GSOEP Big Five Inventory (BFI-S)." *Journal of Research in Personality* 46(3):355–359.

Hastings, Justine S., Christopher A. Neilson, Anely Ramirez, Seth D. Zimmerman and Seth D. Zimmermann. 2016. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." *Economics of Education Review* 51:136–151.

Heckman, James J., John E. Humphries, Gregory Veramendi and Sergio Urzúa. 2014. "Education, Health and Wages." *National Bureau of Economic Research*.

Huntington-Klein, Nick. 2015*a*. "Educational Choice: Subjective Data and Consumption Value." *WP*.

Huntington-Klein, Nick. 2015*b*. "Subjective and projected returns to education." *Journal of Economic Behavior & Organization* 117:10–25.

Huntington-Klein, Nick. 2016. ""(Un)informed College and Major Choice": Verification in an Alternate Setting." *Economics of Education Review* 53:159–163.

Hussey, Andrew J. and Omari H. Swinton. 2011. "Estimating the ex ante Expected Returns to College." *American Economic Review: Papers and Proceedings* 101(3):598–602.

Jaik, Katharina and Stefan C. Wolter. 2016. "The Influence of Locus of Control on Delaying Educational Decisions." *IZA Discussion Paper No. 10191*.

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125(2):515–548.

Keller, Tamás and Guido Neidhöfer. 2014. "Who Dares, Wins? A sibling analysis of tertiary education transition in Germany." *SOEPpapers 713*.

Light, Audrey and Wayne Strayer. 2000. "Determinants of College Completion: School Quality or Student Ability?" *Journal of Human Resources* 35(2):289–308.

Manski, Charles F. 1989. "Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon." *Economics of Education Review* 8(4):305–312.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72(5):1329–1376.

Milla, Joniada. 2014. "Tertiary Education Choices under Uncertainty." *Working Paper*.

Obama, Barack. 2014. "State of the Union Address.".
http://www.whitehouse.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address

Oreopoulos, Philip and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1):159–184.

Papke, Leslie E. and Jeffrey M. Wooldridge. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics* 11(6):619–632.

Papke, Leslie E. and Jeffrey M. Wooldridge. 2008. "Panel data methods for fractional response variables with an application to test pass rates." *Journal of Econometrics* 145(1-2):121–133.

Piatek, Rèmi and Pinger, Pia R. 2015. "Maintaining (Locus of) Control? Data Combination for the Identification and Inference of Factor Structure Models." *Journal of Applied Econometrics* 31:734–755.

Pinger, Pia R. 2015. "Understanding the Mechanisms behind Intergenerational Effects of Economic Distress." *Working Paper*.

Rotter, Julian B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs* 80(1):1–28.

Stange, Kevin E. 2012. "An Empirical Investigation of the Option Value of College Enrollment." *American Economic Journal: Applied Economics* 4(1):49–84.

Statistisches Bundesamt. 2011. Die Neue Berufsbildungsstatistik. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2013. Berufsbildung auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Statistisches Bundesamt. 2014. Schulen auf einen Blick. Technical report Statistisches Bundesamt, Wiesbaden.

Stinebrickner, Ralph and Todd Stinebrickner. 2012. "Learning about academic ability and the college dropout decision." *Journal of Labor Economics* 30(4):707–748.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*a*. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81(1):426–472.

Stinebrickner, Ralph and Todd Stinebrickner. 2014*b*. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32(3):601–644.

Taber, Christopher R. 2000. "Semiparametric identification and heterogeneity in discrete choice dynamic programming models." *Journal of Econometrics* 96(2):201–229.

Taber, Christopher R. 2001. "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* 68(3):665–691.

Tolsma, Jochem, Ariana Need and Uulkje De Jong. 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.

Trachter, Nicholas. 2015. "Stepping stone and option value in a model of postsecondary education." *Quantitative Economics* 6(1):223–256.

Turner, Sarah E. 2004. Going to College and Finishing College Explaining Different Educational Outcomes. In *College Decisions: How Students Actually Make Them and How They Could*, ed. Caroline Hoxby. University of Chicago Press for NBER.

Venti, Steven F. and David A. Wise. 1983. "Individual attributes and self-selection of higher education: College attendance versus college completion." *Journal of Public Economics* 21(1):1–32.

Wiswall, Matthew and Basit Zafar. 2015*a*. "Determinants of college major choice: Identification using an information experiment." *Review of Economic Studies* 82(2):791–824.

Wiswall, Matthew and Basit Zafar. 2015*b*. "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9(2):117–169.

Wölfel, Oliver and Guido Heineck. 2012. "Parental risk attitudes and children's secondary school track choice." *Economics of Education Review* 31(5):727–743.

Zachary, Bleemer and Basit Zafar. 2015. "Intended College Attendance: Evidence from an Experiment on College Returns and Cost." *IZA DP No. 9445*.

Zafar, Basit. 2011*a*. "Can Subjective Expectations Data be used in Choice Models? Evidence on Cognitive Biases." *Journal of Applied Econometrics* 26(3):520–544.

Zafar, Basit. 2011*b*. "How Do College Students Form Expectations?" *Journal of Labor Economics* 29(2):301–348.

A Additional derivations

A.1 Altonji-Elder-Taber with continuous explanatory variable

We estimate models (1) and (5) jointly, imposing the following structure on the unobservables

$$(\upsilon_i, \epsilon_i) \sim \Phi_2(0, 0, 1, 1, \rho), \tag{7}$$

where $\Phi_2(\cdot)$ denotes the bivariate normal distribution, and its arguments are the two errors' means, variances, and their correlation. In other words, we estimate probit models for all outcomes d_i with p_i as a normal endogenous explanatory variable [denoted probit eev hereafter].

The corresponding log-likelihood is given by

$$\ln L(d_i, p_i; x_i, \alpha, \beta^d, \beta^p, \rho) = \sum_{i=1}^{n} \ln \Phi \left[(2d_i - 1) \left(\frac{x_i'\beta^d + \alpha p_i + \rho(p_i - x_i'\beta^p)}{\sqrt{1-\rho^2}} \right) \right] + \ln \phi \left(p_i - x_i'\beta^p \right). \tag{8}$$

For more information, see the discussion in Greene (2012, p747f).

As in AET, we present results for the case when selection on unobservables is equal of the selection on observables, i.e. we estimate the model replacing

$$\rho = \frac{cov(x_i'\beta^d, x_i'\beta^p)}{var(x_i'\beta^d)} \equiv \hat{\rho}^o$$

as a suggestive upper bound.

A.2 Details for Emax in the normal model and error term structure

We use the following definitions from Section 4, where for simplicity we ignore i and t subscripts,

$$\begin{aligned} U_3 &= p\mu_3 + (1-p)\mu_3 + x'\beta_3 + \delta_3'\theta + \varepsilon_3 \equiv z_3 + \varepsilon_3 \,, \\ U_2 &= p\mu_2 + (1-p)\mu_2 + x'\beta_2 + \delta_2'\theta + \varepsilon_2 \equiv z_2 + \varepsilon_2, \end{aligned}$$

and we define

$$\begin{aligned} z_3 - z_2 &\equiv \Delta_2 \,, \\ \nu_2 &= \varepsilon_3 - \varepsilon_2 \,. \end{aligned}$$

When ε_3 and ε_2 are both iid normal we have that

$$\begin{pmatrix} \varepsilon_3 \\ \varepsilon_2 \\ \nu_2 \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 2 \end{bmatrix} \right),$$

and

$$\begin{aligned} E(\varepsilon_3|\nu_2 > -\Delta_2) &= \frac{\sqrt{2}}{2} \cdot \frac{\phi(-\widetilde{\Delta}_2)}{1-\Phi(-\widetilde{\Delta}_2)} = \frac{\sqrt{2}}{2} \cdot \frac{\phi(\widetilde{\Delta}_2)}{\Phi(\widetilde{\Delta}_2)} , \\ E(\varepsilon_2|\nu_2 \leq \Delta_2) &= \frac{\sqrt{2}}{2} \cdot \frac{\phi(-\widetilde{\Delta}_2)}{\Phi(-\widetilde{\Delta}_2)} = \frac{\sqrt{2}}{2} \cdot \frac{\phi(\widetilde{\Delta}_2)}{1-\Phi(\widetilde{\Delta}_2)} , \end{aligned}$$

where $\widetilde{\Delta}_2 \equiv \dfrac{\Delta_2(p)}{\sqrt{2}}$. We can now write,

$$\begin{aligned} E(\max(U_3, z_2)) &= P(U_3 > z_2) \cdot E(U_3|z_3 > z_2) + P(U_3 \leq z_2) \cdot E(U_2|z_3 \leq z_2) \\ &= P(U_3 - U_2 > 0) \cdot (z_3 + E(\varepsilon_3|\nu_2 > -\Delta_2)) \\ &\quad + P(U_3 - U_2 \leq 0) \cdot (z_2 + E(\varepsilon_2|\nu_2 \leq -\Delta_2)) \\ &= \Phi(\widetilde{\Delta}_2) \cdot \big[z_3 + \frac{\sqrt{2}}{2} \cdot \frac{\phi(\widetilde{\Delta}_2)}{\Phi(\widetilde{\Delta}_2)}\big] + (1-\Phi(\widetilde{\Delta}_2)) \cdot \big[z_2 + \frac{\sqrt{2}}{2} \cdot \frac{\phi(\widetilde{\Delta}_2)}{1-\Phi(\widetilde{\Delta}_2)}\big] \\ &= \Phi(\widetilde{\Delta}_2) \cdot z_3 + \frac{\sqrt{2}}{2} \cdot \phi(\widetilde{\Delta}_2) + (1-\Phi(\widetilde{\Delta}_2)) \cdot z_2 + \frac{\sqrt{2}}{2} \cdot \phi(\widetilde{\Delta}_2) \\ &= z_2 + \Phi(\widetilde{\Delta}_2) \cdot (u_3 - u_2) + \sqrt{2} \cdot \phi(\widetilde{\Delta}_2) \\ &= z_2 + \Phi(\widetilde{\Delta}_2) \cdot \widetilde{\Delta}_2 \cdot \sqrt{2} + \sqrt{2} \cdot \phi(\widetilde{\Delta}_2) \\ &= z_2 + \sqrt{2} \cdot \big\{\Phi(\widetilde{\Delta}_2) \cdot \widetilde{\Delta}_2 + \cdot\phi(\widetilde{\Delta}_2)\big\} \end{aligned}$$

The variance-covariance structure of $(\varepsilon_3, \varepsilon_2, \nu_2)$ can be generalized, provided some regressors vary between stage 1 and stage 2 (French and Taber, 2011; Taber, 2000), as in the empirical model presented in Section 4: in this case, one can allow $V(\varepsilon_2) = V(\varepsilon) = \tilde{\sigma}^2$, so that the new covariance matrix is

$$\begin{pmatrix} \varepsilon_3 \\ \varepsilon_2 \\ \nu_2 \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{\sigma}^2 \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 2 \end{bmatrix} \right) ,$$

In Section 4, we parametrize $\sigma^2 = 2\tilde{\sigma}^2$. While the correlation between utilities stemming from the ε's is quite rigid, the inclusion of $\delta_j\theta$ in z_j makes it possible to estimate the variance-covariance structure freely.

B Data

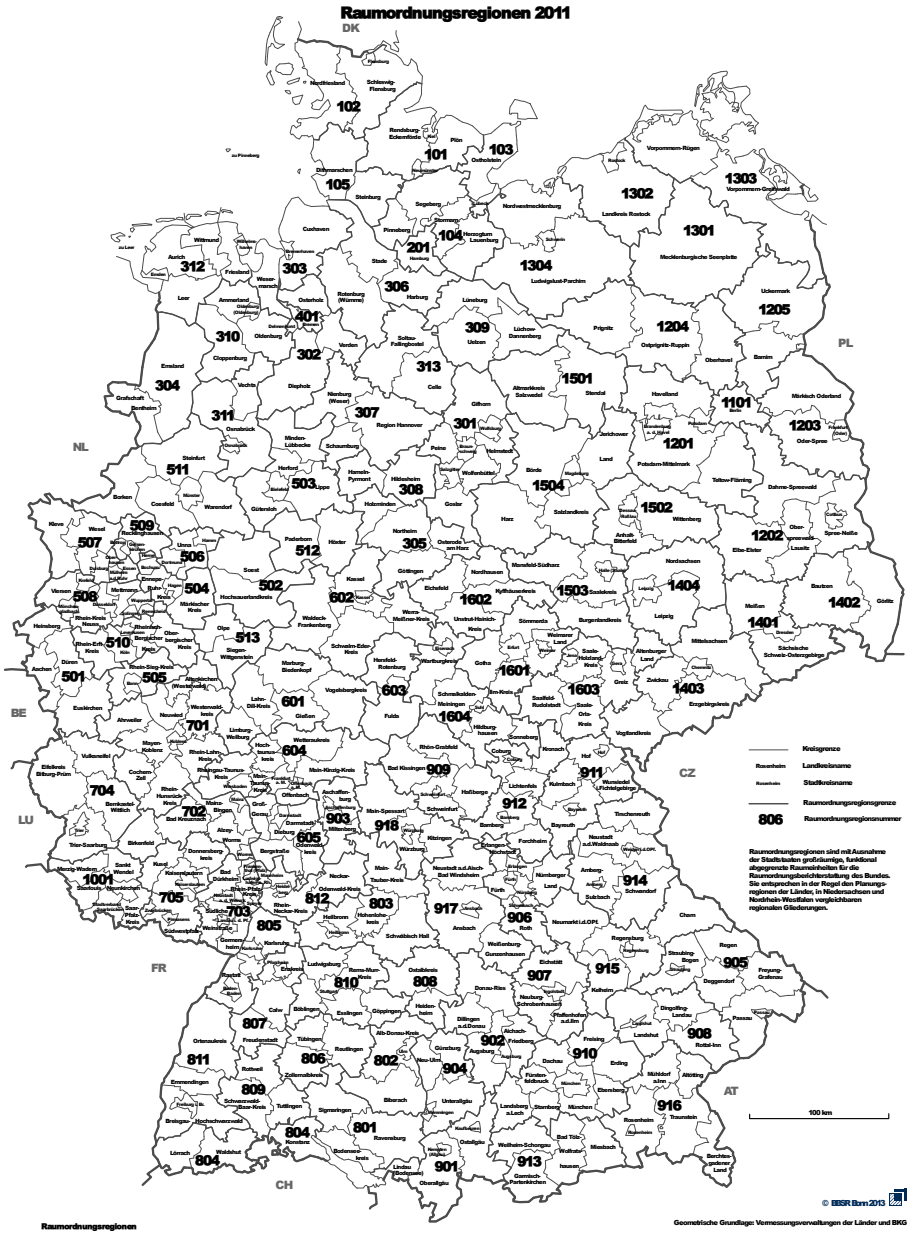


Figure B1: Local labor markets, 96 Raumordnungregionen [Ror]
Source: BBSR (2013)

Table B1: Descriptive statistics by subsample

	(A)	(B)	(B')	(C)	(C')
p	0.776	0.772	0.781	0.769	0.778
	(0.198)	(0.201)	(0.192)	(0.197)	(0.190)
GPA (std)	0.000	-0.018	0.012	0.041	0.077
	(1.000)	(1.019)	(1.009)	(1.024)	(1.016)
Rec: Lowest Track (yes/no)	0.132	0.135	0.134	0.109	0.106
	(0.338)	(0.342)	(0.341)	(0.311)	(0.308)
Rec: Intermediate Track (yes/no)	0.259	0.272	0.270	0.247	0.249
	(0.438)	(0.445)	(0.444)	(0.431)	(0.433)
Rec: High school (yes/no)	0.402	0.347	0.360	0.384	0.400
	(0.490)	(0.476)	(0.480)	(0.487)	(0.490)
In high school (yes/no)	0.411	0.359	0.371	0.415	0.432
	(0.492)	(0.480)	(0.483)	(0.493)	(0.496)
Locus of control (std)	0.000	0.028	0.046	0.055	0.063
	(0.952)	(0.934)	(0.920)	(0.915)	(0.907)
Risk attitudes (std)	0.000	-0.088	-0.087	-0.153	-0.149
	(0.957)	(0.997)	(0.995)	(0.975)	(0.976)
Openness (std)	0.000	-0.014	0.010	0.001	0.027
	(0.946)	(0.978)	(0.971)	(1.014)	(1.014)
Agreeableness (std)	-0.000	-0.007	0.006	-0.011	0.006
	(0.947)	(0.986)	(0.982)	(0.995)	(1.002)
Extraversion (std)	0.000	-0.009	0.009	-0.002	0.023
	(0.946)	(0.953)	(0.949)	(0.963)	(0.964)
Neuroticism (std)	-0.000	-0.015	-0.030	-0.027	-0.046
	(0.946)	(0.979)	(0.974)	(1.019)	(1.021)
Conscientiousness (std)	0.000	0.112	0.132	0.176	0.194
	(0.946)	(0.952)	(0.945)	(0.961)	(0.954)
Female (yes/no)	0.504	0.506	0.508	0.496	0.499
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
Nr. siblings	1.570	1.639	1.630	1.633	1.622
	(1.316)	(1.339)	(1.323)	(1.335)	(1.322)
Second-generation migrant (yes/no)	0.623	0.739	0.731	0.843	0.840
	(0.485)	(0.439)	(0.443)	(0.364)	(0.367)
Parent college-educated (yes/no)	0.283	0.259	0.267	0.292	0.304
	(0.450)	(0.438)	(0.443)	(0.455)	(0.460)
Parent cur. unemployed (yes/no)	0.103	0.124	0.120	0.130	0.129
	(0.304)	(0.329)	(0.325)	(0.336)	(0.335)
Log. net household income	10.377	10.341	10.368	10.374	10.391
	(1.834)	(1.800)	(1.773)	(1.822)	(1.804)
Cyclical youth unemployment (in Ror)	0.074	0.211	0.211	0.267	0.260
	(1.026)	(1.034)	(1.035)	(1.083)	(1.082)
Nr. of apprenticeship positions (in Ror)	98.791	97.920	97.886	97.447	97.362
	(5.279)	(4.988)	(5.039)	(5.034)	(5.067)
Nr. of students (in Ror)	24.095	22.952	23.018	22.395	22.572
	(14.223)	(13.594)	(13.511)	(13.492)	(13.517)
Nr. of high school graduates (in Ror)	26.775	25.137	25.188	24.491	24.582
	(6.673)	(5.514)	(5.520)	(5.155)	(5.158)
Nr. of Universities (in Ror)	10.585	10.333	10.321	10.208	10.225
	(9.938)	(9.850)	(9.849)	(9.651)	(9.637)
N	3'610	2'116	1'919	1'372	1'255

Note: Table presents sample means and standard deviations in brackets in total and by subsample considered in Table 3, and B1 which shows the results of Table 3 conditional on intentions.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table B2: Variable definitions

Variables	Description	Age	Missings
Core variables	*Missing values in the core variables are dropped from the estimation sample.*		
p	Subjective completion belief is a elicited measure, it ranges from 0 to 1, in 0.1 steps.	17	52
GPA (std)	Average of German and Math grades, standardized over the sample population, as a robustness check we additionally standardize within educational track (cf. Table D5).	17	43
Educational outcomes:	From the longitudinal information we assess whether the student has started/completed a respective educational track.	17-31	
$d \in \{0, 1, 2, 3\}$	Intentions/Start/Complete, disaggregated by the tracks: drop out, apprenticeship, tertiary apprenticeship (high school and apprenticeship), and university (includes all higher learning institutions).		
$d_1 \in \{0, 1, 2\}$	First stage in structural model: drop out, apprenticeship, and high school.		
$d_2 \in \{0, 1\}$	Second stage in structural model: tertiary apprenticeship and university.		
Start apprenticeship	Not used in the analysis, all individuals that stared before are dropped from the estimation sample.	17	487
Still in school	Used in intention regressions, but dropped in the investment/completion analysis.	17	1'420
Academic variables			
Recommendations:	To visit a secondary-school track teachers evaluate the students (age the age of 10), the base category is no recommendation, three indicators for Lowest Track (yes/no), Intermediate Track (yes/no), and High school (yes/no)	17	249
In high school (yes/no)	An indicator whether the student is currently in high school when answering the youth questionnaire.	17	105
Personality variables	*We standardize the personality variables to mean 0 and standard deviation 1.*		
Locus of control (std)	First principal component of 10 questions, of which three are reversed. Change in questionnaire adjusted as in Piatek and Pinger (2015).	17	337
Risk attitudes (std)	Assessed by a single question, ranging from 1-10.	17	306
Openness (std)	First principal component of 3 questions.	17	381
Agreeableness (std)	First principal component of 3 questions, of which one is reversed.	17	375
Extraversion (std)	First principal component of 3 questions, of which one is reversed.	17	378
Neuroticism (std)	First principal component of 3 questions, of which one is reversed.	17	378
Conscientiousness (std)	First principal component of 3 questions, of which one is reversed.	17	381
Individual and family characteristics	*Parental information, based on parents' questionnaires, are merged with the children's information.*		
Female (yes/no)	An indicator whether the individual is female.	17	
Nr. of siblings	Count of the number of siblings.	17	179
Second-generation migrant (yes/no)	An indicator whether the individual's parents are born in a foreign country, if information is missing recoded as second-generation migrant.	17	2'029
Parent college-educated (yes/no)	An indicator whether the individual has at least one college educated parent.	17	43

Parent cur. unemployed (yes/no)	An indicator whether the individual has at least one currently unemployed parent.	17	152
Log. net household income	Log of household pre-governmental income imputed by SOEP (0 income is treated as missing)	17	70
Fixed effects			
Year	Year of answering youth questionnaire, which is roughly identical to year of birth	17	
Region	Five regions based on federal states which are the level of educational-jurisdiction, cf. footnote 20	17	109
Regional labor market information	*Information from INKAR 2012/Statistical agency, merged onto the students residence with 17 and lagged by one year. Some are twice assessed for the estimation of the structural model, based on residence with 17 to avoid endogeneity due to moving (there are no missings as the location is always known at 179.*		
Cyclical youth unemployment	Cyclical component of local youth unemployment, extracted by HP-filter.	16/18	
Nr. of apprenticeship positions	Number of apprenticeship positions by all potential apprentices times 100.	16/18	
Nr. of students	Number of students enrolled in higher learning institutions by all residents in the age group times 1000.	16/18	
Nr. of high school graduates	Number of students with a high school degree in the region over all school-leavers times 1000.	16/18	
Nr. of universities	Count of higher learning institutions in the Ror, due to minimal variation over time we keep it constant.	16	

Note: Table presents variable descriptions and missing values for the baseline sample. All available individuals add up to 4'745, which then reduce to 40. The remaining missings are conditional on the estimation sample. All variables besides core variables are included in the estimation along with missing value indicators. More information on the regional indicators can be found under `http://www.inkar.de`

C Additional results

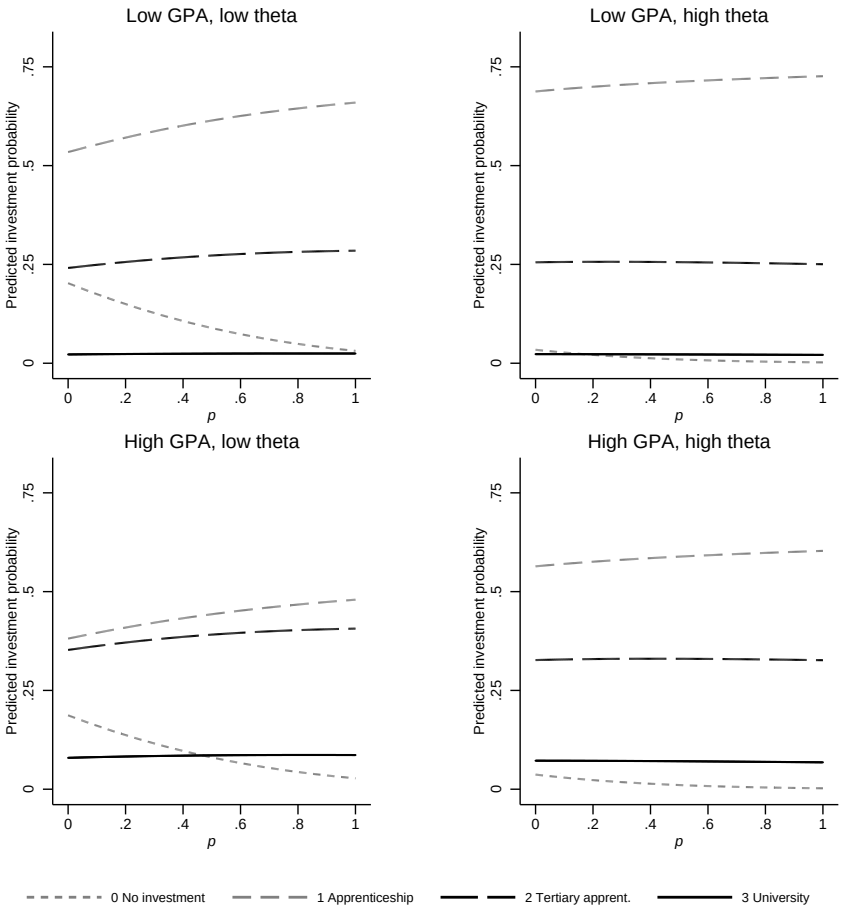
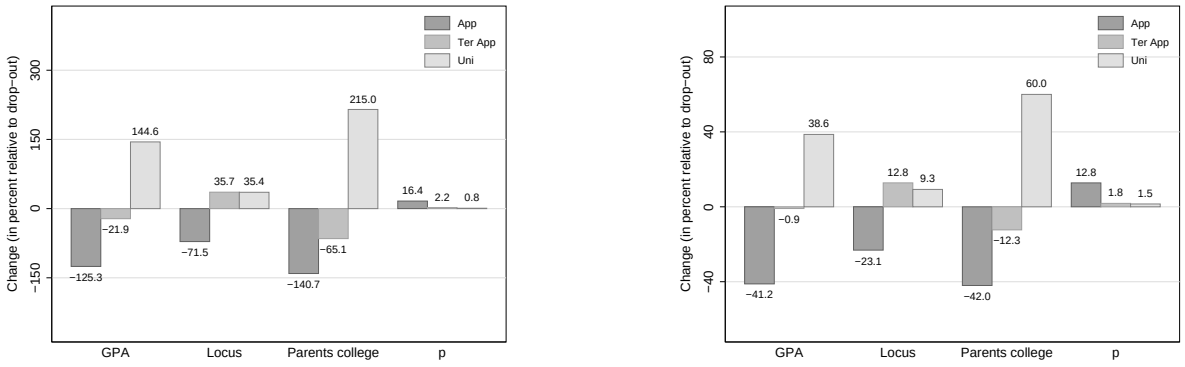


Figure C1: Robustness: The role of academic ability and unobserved heterogeneity

Notes: Predicted probabilities constructed using estimates from Column (8) of Table 4, evaluated at $x'\beta_j = 0$. High GPA and θ are evaluated at the 75th percentiles; low GPA and θ, at the 25th percentile.
Source: SOEP 2000-2013, INKAR 2012, own calculations.



(A) Change in covariates

(B) Change in covariates

Figure C2: Comparison of interventions

Notes: All predictions constructed using estimates from Column (8) of Table 4. Figures (A) "all students": averages over entire sample, θ_i integrated out with mean equal 0. Figures (B) "low ability students": averages over individuals with GPA$\leq$ 0.5, with mean unobserved skills equal to -1. Figures (A), (B): predicted change, in %, in the probability invest in either track for a one standard deviation increase in GPA, locus of control, p, and for a unit increase in parents college.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table C1: Decomposition of predicted changes to "no investment" into direct and total effect

	All				Low ability			
	Direct	s.e.	Total	s.e.	Direct	s.e.	Total	s.e.
p	-19.438	0.073			-16.069	0.061		
GPA (std)	2.605	0.047	-4.847	0.120	3.502	0.048	-3.051	0.117
In high school (yes/no)	-10.500	0.191	-12.424	0.218	-3.661	0.201	-2.769	0.228
Locus of control (std)	0.411	0.026	-7.209	0.115	1.114	0.027	-3.821	0.111
Risk attitudes (std)	-4.616	0.015	-8.160	0.114	-3.770	0.012	-6.271	0.109
Openness (std)	16.004	0.053	12.250	0.121	12.895	0.042	11.033	0.115
Agreeableness (std)	2.801	0.009	-1.529	0.112	2.280	0.008	-0.938	0.108
Extraversion (std)	26.388	0.074	20.190	0.127	19.415	0.053	14.905	0.117
Neuroticism (std)	9.351	0.029	5.977	0.114	7.369	0.021	5.248	0.110
Conscientiousness (std)	-18.683	0.065	-24.913	0.136	-15.658	0.053	-20.225	0.125
Female (yes/no)	-33.016	0.106	-35.338	0.155	-26.934	0.083	-27.377	0.134
No. siblings	5.395	0.017	2.137	0.112	4.298	0.013	2.696	0.109
Second-gen migrant (yes/no)	-52.277	0.179	-51.220	0.200	-44.480	0.145	-38.738	0.160
Parents college (yes/no)	-9.123	0.055	-14.879	0.127	-5.593	0.054	-9.142	0.121
Parent unemployed (yes/no)	57.720	0.153	54.739	0.185	40.022	0.109	41.012	0.155
Log net HH income (in Ror)	-7.085	0.022	-11.280	0.116	-5.714	0.016	-8.524	0.110
Cyclical youth unemployment (in Ror)	-2.659	0.010	-7.113	0.113	-2.214	0.008	-6.219	0.109
No. apprenticeship positions (in Ror)	-11.259	0.037	-14.534	0.119	-9.240	0.030	-10.463	0.112
No. high school graduates (in Ror)	23.283	0.067	20.061	0.128	17.484	0.048	14.753	0.117

Note: Cells contain average predicted changes in percent (and their standard errors in columns "s.e.") of the probability to fail to invest in any post-secondary education, for selected variables. Estimates are from the sequential model with unobserved heterogeneity, presented in Column (8) of Table 4. Columns labelled "Direct" report average predicted changes ceteris paribus. Columns labelled "Total" compute to the ceteris paribus effect the indirect effect through the variables effect on p. Panel "All" reports average effects in the population. Panel "Low ability" reports average effects in the low ability subpopulation as defined in Figure 5 in the main text.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table C2: Disaggregated intentions-to-invest, multinominal probit regressions

Dependent variables: Jointly estimated intention to invest in apprenticeship, tertiary apprenticeship, and university.				
	(1)	(2)	(3)	(4)
Apprenticeship, $d = 1$				
p	1.045	1.136	1.032	1.056
	(0.206)	(0.214)	(0.219)	(0.222)
	[0.021(0.029)]	[0.081(0.029)]	[0.081(0.030)]	[0.089(0.030)]
Tertiary Apprenticeship, $d = 2$				
p	1.094	0.907	0.817	0.791
	(0.210)	(0.221)	(0.227)	(0.230)
	[-0.011(0.034)]	[-0.003(0.035)]	[0.011(0.036)]	[0.010(0.036)]
University, $d = 3$				
p	1.557	1.073	0.834	0.779
	(0.217)	(0.227)	(0.236)	(0.239)
	[0.141(0.035)]	[0.052(0.035)]	[0.020(0.036)]	[0.011(0.036)]
N	3'610	3'610	3'610	3'610
F (pval)	54.900 (0.000)	36.009 (0.000)	25.779 (0.000)	25.511 (0.000)
Fixed effects	✓	✓	✓	✓
Academic		✓	✓	✓
Personality			✓	✓
Family Background				✓
Labor market				✓

Note: Table presents, multinominal probit regressions of the educational intention-to-invest: drop out, apprenticeship, tertiary apprenticeship, and university on subjective beliefs and varying sets of covariates in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. Robust standard errors in round, average marginal effect along (with standard errors) in squared (round) brackets. The Likelihood Ratio (LR)-statistic measures the significance of p across equations.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table C3: Disaggregated actual investment and completion, conditional on intentions, probit regressions

Dependent variables: Separately estimated actual invest in and completion of apprenticeship, tertiary apprenticeship, and university.								
	Actual investment				Actual completion			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Apprenticeship ($d_1 = 1$)								
p	0.902	1.055	0.987	1.034	0.794	0.737	0.616	0.563
	(0.256)	(0.266)	(0.269)	(0.281)	(0.271)	(0.273)	(0.276)	(0.277)
	[0.133]	[0.151]	[0.135]	[0.092]	[0.292]	[0.271]	[0.226]	[0.206]
R^2_n	0.033	0.044	0.072	0.170	0.051	0.060	0.072	0.102
$R^2_n(p)$	0.054	0.071	0.094	0.191	0.064	0.071	0.079	0.108
	$N = 883$, $\bar{d}_1 = 0.95$, $\bar{p} = 0.77$, $SD(p) = 0.22$				$N = 502$, $\bar{d}_1 = 0.67$, $\bar{p} = 0.76$, $SD(p) = 0.22$			
(B) Tertiary apprenticeship ($d_2 = 1$)								
p	0.362	0.455	0.591	0.684	0.409	0.358	0.272	0.522
	(0.368)	(0.379)	(0.415)	(0.417)	(0.549)	(0.564)	(0.593)	(0.626)
	[0.120]	[0.150]	[0.192]	[0.218]	[0.098]	[0.089]	[0.067]	[0.101]
R^2_n	0.079	0.084	0.110	0.138	0.140	0.139	0.155	0.226
$R^2_n(p)$	0.081	0.087	0.113	0.142	0.142	0.141	0.155	0.228
Sample:	$N = 456$, $\bar{d}_2 = 0.97$, $\bar{p} = 0.78$, $SD(p) = 0.18$				$N = 314$, $\bar{d}_2 = 0.56$, $\bar{p} = 0.78$, $SD(p) = 0.18$			
(C) University ($d_3 = 1$)								
p	0.789	0.431	0.459	0.219	0.980	0.665	0.927	0.671
	(0.367)	(0.388)	(0.412)	(0.416)	(0.487)	(0.512)	(0.544)	(0.549)
	[0.282]	[0.153]	[0.161]	[0.075]	[0.253]	[0.172]	[0.232]	0.158
R^2_n	0.140	0.186	0.210	0.258	0.177	0.194	0.223	0.251
$R^2_n(p)$	0.146	0.188	0.211	0.259	0.186	0.198	0.229	0.254
Sample:	$N = 580$, $\bar{d}_3 = 0.98$, $\bar{p} = 0.80$, $SD(p) = 0.15$				$N = 428$, $\bar{d}_3 = 0.40$, $\bar{p} = 0.80$, $SD(p) = 0.16$			
Fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓		✓	✓	✓
Personality			✓	✓			✓	✓
Family Background				✓				✓
Labor market				✓				✓

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit regressions of investment (1)-(4) and completion (5)-(8) on subjective completion beliefs and varying sets of covariate, in (1/5) on in high school, region and time fixed effects, (2/6) adds academic, (3/7) adds personality, (4/8) family background, individual, and local labor market characteristics. We present McFadden's pseudo-R^2 and sample statistics for the varying subsamples. For some regressions the numbers of observations are slightly reduced.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

D Robustness

Table D1: Determinants of subjective completion beliefs, fractional response regressions

Dependent variable: Subjective completion belief with unconditional mean 0.776 and standard deviation 0.198				
	(1)	(2)	(3)	(4)
GPA (std)	0.037	0.029	0.029	0.029
	(0.003)	(0.003)	(0.003)	(0.003)
Rec: Lowest Track (yes/no)	0.026	0.024	0.026	0.026
	(0.013)	(0.013)	(0.013)	(0.013)
Rec: Intermediate Track (yes/no)	0.060	0.054	0.052	0.051
	(0.010)	(0.010)	(0.010)	(0.010)
Rec: High school (yes/no)	0.049	0.040	0.036	0.036
	(0.010)	(0.010)	(0.010)	(0.010)
In high school (yes/no)	0.003	0.001	-0.006	-0.006
	(0.008)	(0.008)	(0.008)	(0.008)
Locus of control (std)		0.024	0.023	0.023
		(0.004)	(0.004)	(0.004)
Risk attitudes (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Openness (std)		0.005	0.004	0.004
		(0.004)	(0.004)	(0.004)
Agreeableness (std)		0.006	0.007	0.007
		(0.004)	(0.004)	(0.004)
Extraversion (std)		0.016	0.017	0.017
		(0.004)	(0.004)	(0.004)
Neuroticism (std)		-0.002	-0.000	-0.000
		(0.004)	(0.004)	(0.004)
Conscientiousness (std)		0.031	0.032	0.033
		(0.004)	(0.004)	(0.004)
Female (yes/no)			-0.010	-0.009
			(0.007)	(0.007)
Nr. siblings			-0.003	-0.003
			(0.002)	(0.002)
Second-generation migrant (yes/no)			-0.008	-0.008
			(0.013)	(0.013)
Parent college-educated (yes/no)			0.009	0.008
			(0.007)	(0.007)
Parent cur. unemployed (yes/no)			0.003	0.003
			(0.012)	(0.012)
Log. net household income			0.006	0.006
			(0.002)	(0.002)
N	3'610	3'610	3'610	3'610
R_n^2	0.065	0.128	0.134	0.134
Fixed effects	✓	✓	✓	✓
Academic	✓	✓	✓	✓
F (p-value)	185.410 (0.000)	115.637 (0.000)	102.904 (0.000)	102.956 (0.000)
Personality		✓	✓	✓
F (p-value)		218.885 (0.000)	216.912 (0.000)	218.254 (0.000)
Family Background			✓	✓
F (p-value)			17.503 (0.014)	16.847 (0.018)
Labor market				✓
F (p-value)				1.819 (0.874)

Note: The Table presents Bernoulli pseudo-maximum likelihood with probit conditional expectation function, as proposed by Papke and Wooldridge (1996, 2008). We report marginal effects and robust standard errors in round brackets, our goodness of fit measure is a nonlinear R-squared measure and is calculated as the squared correlation coefficient between the estimated conditional expectation and the observed subjective beliefs: $R_n^2 = corr(\hat{p}, p)^2$, where $\hat{p} = \Phi(x'\hat{\beta})$ due to the probit specification. The regressions of subjective beliefs are presented on varying sets of covariates, in (1) only on academic region and time fixed effects (coefficients not presented), (2) adds personality, (3) family background and individual characteristics, and (4) local labor market characteristics. We present the unconditional mean $\bar{p}$ and standard deviation $SD(p)$ of the dependent variable.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table D2: Robustness: Conditional on intentions and subjective investment beliefs

Dependent variable: Indicator variables for actual investment, and completion.							
	probit				probit eev		
					$\rho = \hat{\rho}^o$	$\rho = .1$	$\rho = .3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(B') Actual investment, conditional on intentions							
p	0.901	0.845	0.814	0.713	0.702	0.609	0.380
	(0.256)	(0.262)	(0.268)	(0.275)	(0.275)	(0.274)	(0.262)
	[0.058]	[0.053]	[0.048]	[0.032]			
R_n^2	0.085	0.095	0.111	0.201			
$R_n^2(p)$	0.104	0.111	0.124	0.210			
Sample: $N = 1'919$, $\bar{d} = 0.96$, $\bar{p} = 0.78$, $SD(p) = 0.19$, $\hat{\rho}^o = 0.010$							
(B") Actual investment, conditional on investment beliefs							
p	0.997	0.915	0.902	0.862	0.610	0.553	0.326
	(0.223)	(0.228)	(0.239)	(0.249)	(0.296)	(0.295)	(0.283)
	[0.069]	[0.062]	[0.056]	[0.044]			
R_n^2	0.087	0.100	0.120	0.182			
$R_n^2(p)$	0.113	0.121	0.138	0.197			
Sample: $N = 2'116$, $\bar{d} = 0.96$, $\bar{p} = 0.77$, $SD(p) = 0.20$, $\hat{\rho}^o = 0.045$							
(C') Actual completion, conditional on intentions							
p	0.467	0.478	0.466	0.456	0.457	0.353	0.135
	(0.198)	(0.202)	(0.208)	(0.211)	(0.211)	(0.210)	(0.201)
	[0.185]	[0.189]	[0.185]	[0.180]			
R_n^2	0.095	0.098	0.106	0.125			
$R_n^2(p)$	0.099	0.101	0.109	0.128			
Sample: $N = 1'244$, $\bar{d} = 0.55$, $\bar{p} = 0.78$, $SD(p) = 0.19$, $\hat{\rho}^o(se) = -0.001$							
(C") Actual completion, conditional on investment beliefs							
p	0.434	0.410	0.387	0.373	0.194	0.113	-0.096
	(0.181)	(0.185)	(0.190)	(0.192)	(0.211)	(0.210)	(0.201)
	[0.172]	[0.162]	[0.147]	[0.140]			
R_n^2	0.089	0.093	0.102	0.121			
$R_n^2(p)$	0.092	0.096	0.104	0.123			
Sample: $N = 1'372$, $\bar{d} = 0.54$, $\bar{p} = 0.77$, $SD(p) = 0.20$, $\hat{\rho}^o = 0.019$							
Fixed effects	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓	✓	✓	✓
Personality			✓	✓	✓	✓	✓
Family Background				✓	✓	✓	✓
Labor market				✓	✓	✓	✓

Note: Table presents robustness from the main results in Table 3, as in presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics. We present the regressions, restricting the sample to those with positive intentions in Panel B' and C'. Alternatively, we include the subjective investment probability as an additional covariate in all regressions of Panel B" and C"

Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table D3: Robustness: dichotomizing subjective beliefs ($p \geq 0.70$)

Dependent variable: Indicator variables for educational intentions, actual investment, and completion.								
	probit				bivariate probit			
					$\rho = .05$	$\rho = .1$	$\rho = .2$	$\rho = .3$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intentions								
p	0.381	0.321	0.276	0.266	0.181	0.095	-0.075	-0.243
	(0.066)	(0.069)	(0.070)	(0.070)	(0.070)	(0.070)	(0.069)	(0.068)
	[0.068]	[0.055]	[0.046]	[0.044]				
R_n^2	0.029	0.040	0.048	0.057				
$R_n^2(p)$	0.043	0.049	0.055	0.063				
Sample: $N = 3'610$, $\bar{d} = 0.91$, $\bar{p} = 0.79$, $SD(p) = 0.41$								
(B) Actual investment								
p	0.491	0.453	0.439	0.431	0.345	0.259	0.088	-0.083
	(0.108)	(0.111)	(0.113)	(0.119)	(0.118)	(0.118)	(0.116)	(0.114)
	[0.045]	[0.039]	[0.035]	[0.029]				
R_n^2	0.087	0.100	0.120	0.182				
$R_n^2(p)$	0.112	0.120	0.137	0.197				
Sample: $N = 2'116$, $\bar{d} = 0.96$, $\bar{p} = 0.79$, $SD(p) = 0.41$								
(B') Actual investment, conditional on intentions								
p	0.476	0.453	0.441	0.440	0.353	0.266	0.094	-0.077
	(0.120)	(0.122)	(0.123)	(0.129)	(0.129)	(0.128)	(0.127)	(0.124)
	[0.040]	[0.037]	[0.034]	[0.025]				
R_n^2	0.085	0.095	0.111	0.201				
$R_n^2(p)$	0.108	0.114	0.128	0.216				
Sample: $N = 1'919$, $\bar{d} = 0.96$, $\bar{p} = 0.81$, $SD(p) = 0.40$								
(C) Actual completion								
p	0.236	0.225	0.219	0.217	0.130	0.043	-0.132	-0.307
	(0.089)	(0.091)	(0.093)	(0.094)	(0.094)	(0.094)	(0.093)	(0.091)
	[0.094]	[0.089]	[0.087]	[0.086]				
R_n^2	0.089	0.093	0.102	0.121				
$R_n^2(p)$	0.093	0.097	0.105	0.124				
Sample: $N = 1'372$, $\bar{d} = 0.54$, $\bar{p} = 0.79$, $SD(p) = 0.41$								
(C') Actual completion, conditional on intentions								
p	0.266	0.271	0.273	0.269	0.181	0.093	-0.083	-0.261
	(0.097)	(0.099)	(0.101)	(0.102)	(0.102)	(0.101)	(0.100)	(0.099)
	[0.106]	[0.108]	[0.108]	[0.107]				
R_n^2	0.095	0.098	0.106	0.125				
$R_n^2(p)$	0.100	0.103	0.111	0.129				
Sample: $N = 1'244$, $\bar{d} = 0.55$, $\bar{p} = 0.81$, $SD(p) = 0.39$								
Fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓	✓	✓	✓	✓
Personality			✓	✓	✓	✓	✓	✓
Family Background				✓	✓	✓	✓	✓
Labor market				✓	✓	✓	✓	✓

Note: Table presents robustness of the main Table 3, where instead of the continuous p we use a dummy for $p \leq 70$. The Table is analogous to Table 3. In the bivariate probit regressions we restrict the correlation between the errors to be 0.05, 0.1, 0.2, 0.3.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table D4: Robustness: separate regressions by high school attendance

Dependent variable: Indicator variables for educational intentions, actual investment, and completion.								
	Not in high school				In high school			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intentions								
p	0.989	0.880	0.789	0.788	0.655	0.446	0.425	0.454
	(0.160)	(0.164)	(0.167)	(0.168)	(0.303)	(0.311)	(0.329)	(0.338)
	[0.179]	[0.158]	[0.140]	[0.137]	[0.071]	[0.048]	[0.043]	[0.043]
R_n^2	0.015	0.030	0.043	0.050	0.030	0.042	0.061	0.088
$R_n^2(p)$	0.043	0.051	0.058	0.066	0.036	0.044	0.063	0.090
N	2'125	2'125	2'125	2'125	1'485	1476	1476	1476
(B) Actual investment								
p	1.018	0.993	0.978	0.975	1.122	0.557	0.006	-0.679
	(0.241)	(0.249)	(0.259)	(0.277)	(0.622)	(0.605)	(0.711)	(0.872)
	[0.097]	[0.092]	[0.085]	[0.068]	[0.036]	[0.010]	[0.000]	[0.000]
R_n^2	0.065	0.080	0.102	0.181	0.078	0.187	0.322	0.439
$R_n^2(p)$	0.095	0.106	0.126	0.202	0.099	0.191	0.322	0.441
N	1'356	1'356	1'356	1'356	584	582	582	582
(B') Actual investment, conditional on intentions								
p	0.943	0.979	0.928	0.877	0.710	0.088	-0.678	-3.016
	(0.277)	(0.285)	(0.285)	(0.310)	(0.653)	(0.651)	(0.835)	(1.212)
	[0.080]	[0.080]	[0.071]	[0.043]	[0.024]	[0.002]	[-0.002]	[0.000]
R_n^2	0.078	0.090	0.113	0.244	0.071	0.183	0.324	0.475
$R_n^2(p)$	0.100	0.113	0.131	0.257	0.078	0.183	0.327	0.496
N	1'207	1'207	1'207	1'207	546	544	544	544
(C) Actual completion								
p	0.534	0.501	0.427	0.389	0.247	0.193	0.376	0.399
	(0.212)	(0.215)	(0.219)	(0.224)	(0.347)	(0.359)	(0.389)	(0.399)
	[0.203]	[0.190]	[0.162]	[0.147]	[0.097]	[0.076]	[0.147]	[0.153]
R_n^2	0.053	0.062	0.071	0.102	0.115	0.120	0.145	0.170
$R_n^2(p)$	0.059	0.067	0.074	0.104	0.116	0.120	0.146	0.172
N	802	802	802	802	570	570	570	570
(C') Actual completion, conditional on intentions								
p	0.487	0.500	0.442	0.417	0.460	0.411	0.586	0.629
	(0.237)	(0.241)	(0.246)	(0.252)	(0.362)	(0.374)	(0.402)	(0.413)
	[0.183]	[0.187]	[0.165]	[0.155]	[0.180]	[0.161]	[0.229]	[0.240]
R_n^2	0.059	0.067	0.075	0.104	0.110	0.114	0.134	0.160
$R_n^2(p)$	0.063	0.072	0.078	0.107	0.112	0.115	0.137	0.163
N	709	709	709	709	535	535	535	535
FE	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓		✓	✓	✓
Personality			✓	✓			✓	✓
Family Background				✓				✓
Labor market				✓				✓

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table D5: Robustness: GPA standardized within high school and using federal states fixed effects

Dependent variable: Indicator variables for educational intentions, actual investment, and completion.								
	Federal states, Fixed effects				GPA, std by federal states			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intentions								
p	0.928	0.817	0.723	0.710	0.921	0.811	0.718	0.706
	(0.142)	(0.146)	(0.149)	(0.150)	(0.142)	(0.146)	(0.149)	(0.151)
	[0.141]	[0.123]	[0.108]	[0.104]	[0.140]	[0.122]	[0.107]	[0.103]
R_n^2	0.027	0.038	0.047	0.055	0.029	0.040	0.048	0.056
$R_n^2(p)$	0.048	0.054	0.058	0.066	0.049	0.055	0.059	0.067
Sample: $N = 3'610$, $\bar{d} = 0.91$, $\bar{p} = 0.78$, $SD(p) = 0.20$								
(B) Actual investment								
p	1.033	0.959	0.941	0.882	0.997	0.923	0.910	0.862
	(0.224)	(0.229)	(0.242)	(0.253)	(0.223)	(0.228)	(0.239)	(0.249)
	[0.071]	[0.064]	[0.058]	[0.045]	[0.069]	[0.062]	[0.057]	[0.044]
R_n^2	0.089	0.102	0.121	0.179	0.087	0.100	0.120	0.182
$R_n^2(p)$	0.117	0.124	0.141	0.195	0.113	0.121	0.138	0.197
Sample: $N = 2'116$, $\bar{d} = 0.96$, $\bar{p} = 0.77$, $SD(p) = 0.20$								
(B') Actual investment, conditional on intentions								
p	0.924	0.877	0.839	0.713	0.901	0.858	0.825	0.716
	(0.255)	(0.263)	(0.270)	(0.278)	(0.256)	(0.261)	(0.267)	(0.274)
	[0.058]	[0.054]	[0.048]	[0.032]	[0.058]	[0.054]	[0.048]	[0.032]
R_n^2	0.089	0.098	0.114	0.200	0.085	0.095	0.110	0.201
$R_n^2(p)$	0.108	0.115	0.128	0.209	0.104	0.110	0.124	0.210
Sample: $N = 1'919$, $\bar{d} = 0.96$, $\bar{p} = 0.78$, $SD(p) = 0.19$								
(C) Actual completion								
p	0.408	0.378	0.363	0.350	0.434	0.412	0.390	0.377
	(0.180)	(0.184)	(0.189)	(0.191)	(0.181)	(0.185)	(0.190)	(0.192)
	[0.162]	[0.150]	[0.144]	[0.139]	[0.172]	[0.163]	[0.155]	[0.150]
R_n^2	0.078	0.082	0.091	0.110	0.089	0.093	0.102	0.121
$R_n^2(p)$	0.081	0.085	0.093	0.112	0.092	0.096	0.104	0.123
Sample: $N = 1'372$, $\bar{d} = 0.54$, $\bar{p} = 0.77$, $SD(p) = 0.20$								
(C') Actual completion, conditional on intentions								
p	0.454	0.455	0.456	0.443	0.467	0.479	0.468	0.458
	(0.197)	(0.200)	(0.206)	(0.210)	(0.198)	(0.202)	(0.208)	(0.211)
	[0.180]	[0.180]	[0.180]	[0.175]	[0.185]	[0.190]	[0.185]	[0.181]
R_n^2	0.084	0.087	0.095	0.114	0.095	0.098	0.106	0.125
$R_n^2(p)$	0.088	0.090	0.098	0.117	0.099	0.101	0.109	0.128
Sample: $N = 1'244$, $\bar{d} = 0.55$, $\bar{p} = 0.78$, $SD(p) = 0.19$								
Fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓		✓	✓	✓
Personality			✓	✓			✓	✓
Family Background				✓				✓
Labor market				✓				✓

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.

Table D6: Robustness: GPA standardized within federal states and including a fifth order polynomial in GPA

Dependent variable: Indicator variables for educational intentions, actual investment, and completion.								
	GPA, std by high school attendance				GPA, polynomial			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Intentions								
p	0.921	0.810	0.717	0.704	0.921	0.807	0.715	0.704
	(0.142)	(0.146)	(0.149)	(0.151)	(0.142)	(0.146)	(0.150)	(0.151)
	[0.140]	[0.122]	[0.107]	[0.103]	[0.140]	[0.121]	[0.106]	[0.103]
R_n^2	0.029	0.040	0.048	0.056	0.029	0.041	0.049	0.058
$R_n^2(p)$	0.049	0.055	0.059	0.067	0.049	0.056	0.060	0.068
Sample: $N = 3'610$, $\bar{d} = 0.91$, $\bar{p} = 0.78$, $SD(p) = 0.20$								
(B) Actual investment								
p	0.997	0.918	0.905	0.866	0.997	0.889	0.877	0.832
	(0.223)	(0.228)	(0.240)	(0.249)	(0.223)	(0.227)	(0.239)	(0.250)
	[0.069]	[0.062]	[0.057]	[0.044]	[0.069]	[0.057]	[0.052]	[0.040]
R_n^2	0.087	0.100	0.120	0.182	0.087	0.109	0.129	0.191
$R_n^2(p)$	0.113	0.121	0.138	0.197	0.113	0.128	0.146	0.205
Sample: $N = 2'116$, $\bar{d} = 0.96$, $\bar{p} = 0.77$, $SD(p) = 0.20$								
(B') Actual investment, conditional on intentions								
p	0.901	0.849	0.818	0.717	0.901	0.838	0.810	0.703
	(0.256)	(0.262)	(0.268)	(0.275)	(0.256)	(0.263)	(0.270)	(0.280)
	[0.058]	[0.053]	[0.048]	[0.032]	[0.058]	[0.050]	[0.045]	[0.030]
R_n^2	0.085	0.095	0.110	0.201	0.085	0.102	0.119	0.208
$R_n^2(p)$	0.104	0.110	0.124	0.210	0.104	0.117	0.132	0.216
Sample: $N = 1'919$, $\bar{d} = 0.96$, $\bar{p} = 0.78$, $SD(p) = 0.19$								
(C) Actual completion								
p	0.434	0.410	0.387	0.373	0.434	0.421	0.403	0.392
	(0.181)	(0.184)	(0.190)	(0.192)	(0.181)	(0.185)	(0.191)	(0.193)
	[0.172]	[0.162]	[0.153]	[0.148]	[0.172]	[0.167]	[0.160]	[0.155]
R_n^2	0.089	0.093	0.102	0.121	0.089	0.097	0.105	0.125
$R_n^2(p)$	0.092	0.096	0.104	0.123	0.092	0.099	0.108	0.127
Sample: $N = 1'372$, $\bar{d} = 0.54$, $\bar{p} = 0.77$, $SD(p) = 0.20$								
(C') Actual completion, conditional on intentions								
p	0.467	0.479	0.467	0.456	0.467	0.498	0.488	0.476
	(0.198)	(0.202)	(0.208)	(0.211)	(0.198)	(0.203)	(0.209)	(0.212)
	[0.185]	[0.189]	[0.185]	[0.180]	[0.185]	[0.197]	[0.193]	[0.189]
R_n^2	0.095	0.098	0.106	0.125	0.095	0.102	0.109	0.128
$R_n^2(p)$	0.099	0.101	0.109	0.128	0.099	0.105	0.113	0.131
Sample: $N = 1'244$, $\bar{d} = 0.55$, $\bar{p} = 0.78$, $SD(p) = 0.19$								
FE	✓	✓	✓	✓	✓	✓	✓	✓
Academic		✓	✓	✓		✓	✓	✓
Personality			✓	✓			✓	✓
Family Background				✓				✓
Labor market				✓				✓

Note: Table presents coefficients (robust standard errors in round and average marginal effects in squared brackets), from probit (1)-(4) and probit endogenous explanatory variable (5)-(8) regressions of varying educational outcomes on subjective completion beliefs and varying sets of covariate, in (1) on in high school, region and time fixed effects, (2) adds academic, (3) adds personality, (4) to (8) family background, individual, and local labor market characteristics.
Source: SOEP 2000-2013, INKAR 2012, own calculations.