COLOGNE GRADUATE SCHOOL
in Management, Economics and Social Sciences

Vol 1 | No 7

CGS WORKING PAPER

The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course

An Empirical Test of a Subjective-Expected-Utility Explanation

Dominik Becker (CGS, University of Cologne)

Cologne Graduate School
in Management, Economics
and Social Sciences
Albertus-Magnus-Platz
50923 Köln
www.cgs.uni-koeln.de

University of Cologne

The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course

An Empirical Test of A Subjective-Expected-Utility Explanation

by Dominik Becker

Version of December 15, 2010

Author's Affiliation:

Dominik Becker, M.A.
University of Cologne
Cologne Graduate School in Management, Economics and Social Sciences
Richard-Strauss-Str. 2
50931 Cologne
Germany
mail: dominik.becker@wiso.uni-koeln.de
phone: +49-(0)221-470-1221
fax: +49-(0)221-470-1229

The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course

An Empirical Test of A Subjective-Expected Utility Explanation

The substantial aim of this paper is to integrate the main idea of 'Pygmalion' or self-fulfilling prophecy research (Rosenthal and Jacobson, 1968; Jussim and Harber, 2005) into the general subjective expected utility framework about inequality in educational opportunities (Breen and Goldthorpe, 1997; Esser, 1999). In the theoretical section, a formal model of the impact of self-fulfilling prophecies on educational transitions is developed. In the empirical section, we test this model to predict both students' educational success (in terms of high school graduation) and their university transitions. Since we assume a conditional dependence of these outcomes, we control for sample selection bias (Heckman, 1979). We find that in our operationalization of self-fulfilling prophecies the latter show significant effects on both educational success and university transitions. However, while the results remain stable in case of educational success, we find that the conditional decision problem of university transitions leads to a selection bias for the estimates in the latter case. In a sensitivity analysis we find that only if unobserved heterogeneity would be disturbingly high, it could also affect the stability of self-fulfilling prophecy estimates.

1 Introduction

School surely is the first and by that way also the most important direction point in everybody's life course. Following Schelsky (1957), it is crucial "for everybody's future social security, future social position and the amount of future consumption possibilities" (Schelsky, 1957, 18). The economic literature provides numerous examples for the relationship between schooling and labor market income (e.g. Boissiere et al., 1985; Ashenfelter et al., 1999). Moreover, there is even evidence that in the long run human capital – measured by labor-force quality – may influence nations' productivity and economic growth (Bishop, 1989; Hanushek and Kimko, 2000). However, although the importance of schooling and its quality is undisputed, theories in social science about social inequality in educational opportunities (IEO) still need to be refined.

On the one hand, the theoretical framework that has been provided by social inequality theory based on rational-choice or subjective expected utility (SEU) assumptions surely is powerful. One main strength is that it allows us to distinguish between primary and secondary effects of social inequality, i.e. the difference between effects of socialization and effects of aspirations. Furthermore, SEU theory always implies the formalization of the researcher's assumptions which facilitates both the comparison of different hypotheses and their operationalization into empirical models.

On the other hand, social psychologists have impressively revealed how teachers' expectations can influence students' future performance beyond their (or their parents') mere cost-benefit considerations. This phenomenon has been labeled the 'Pygmalion effect' of self-fulfilling under-estimations and the 'Golem effect' of self-fulfilling over-estimations; (Rosenthal and Jacobson, 1968). Moreover, Pygmalion research showed that the variance of this effect can partially be explained by social background variables (Jussim and Harber, 2005).

The substantial aim of this paper is to integrate the main idea of 'Pygmalion' into the general subjective expected utility framework about IEO. In particular, we will refer to Esser's (1999) extension of the formal IEO model that has been suggested by Breen and Goldthorpe (1997). Furthermore, while many – not all – applications of this model have considered educational transition decisions from primary to secondary school, in this study we will focus on students' probability of achieving a high school degree and on their propensity of beginning academic studies, respectively. The research design will be an extension of Becker's (2003) model which includes controls for selection bias (Heckman, 1979). Additionally, we will perform a sensitivity analysis for all self-fulfilling prophecy indicators to test their robustness against a vector of unobserved covariates (Buis, 2007, 2010).

This paper will be structured as follows: First, the basic assumptions of both the SEU-IEO model and 'Pygmalion' will be discussed. Then we will outline how the implications of the latter require to rebuild the present SEU-IEO model in order to specify the endogeneity of students' subjective expected probability of educational success more adequately. After a short description of the dataset and the variables, a series of stepwise logit models both without and with controls for selection bias will be presented and discussed. These models are amended by the sensitivity analyses for the self-fulfilling prophecy indicators. The paper ends with a conclusion and provides an outlook on potential extensions of the model.

2 Theory and Hypotheses

2.1 Inequality of Educational Opportunities: The Subjected-Expected-Utility Model

One important theoretical concept in the IEO framework is about the differentiation between primary and secondary effects of social inequality (Boudon, 1974). While *primary effects* capture the relationship between social background variables and pupils' academic ability (however the latter will be measured), *secondary effects* of social inequality are defined either as conditions of schools' structure or organization – but mainly as the lower educational aspirations of the students themselves or of their parents (Müller-Benedict, 2007). The argument is that secondary effects of social inequality are still present after having controlled for all primary effects, i.e. given an intelligence score of a certain level, "working class" children will still have lower school achievements because of lower educational aspirations. But why?

Given education as an investment good (Goldthorpe, 1996, p. 494), the chief concern for each family will be to achieve some kind of *intergenerational stability of class positions.* Hence, service-class parents will be more likely than others to encourage their children to attain higher education of some kind. Reversely, for families in less advantaged positions not only less ambitious and less costly educational options would be adequate for the goal of maintaining class stability – but also each *failed* attempt in obtaining higher educational levels is likely to be more serious in its consequences (e.g. in terms of further opportunity costs which have to be shouldered). The main advantage of regarding education as an investment good can be seen in the possibility of applying rational-choice (RC) or subjective expected utility (SEU) functions.

The Breen-Goldthorpe Model Breen and Goldthorpe (1997) provided a formal model in order to account for class differences in educational transitions (see Figure 1a). Let c denote the costs of remaining at school (in terms of both direct costs of educations and earnings forgone), π the subjective likelihood of success if a student continues in education, and P, F and L the value or utility that children or their parents attach to three different academic outcomes. Service-class, working-class and underclass children are denoted by $S*$, $W*$, and $U*$, respectively. According to the two (preliminary) assumptions that i) students' ability is equally distributed among social classes and ii) continuing in education is cost-less, the probabilities of service-class children to remain in service class conditional on the chosen school track are given by

$$p_{is} = \frac{\pi_i \alpha + (1-\pi_i)\beta_1}{\pi_i \alpha + (1-\pi_i)\beta_1 + \gamma_1} \tag{1}$$

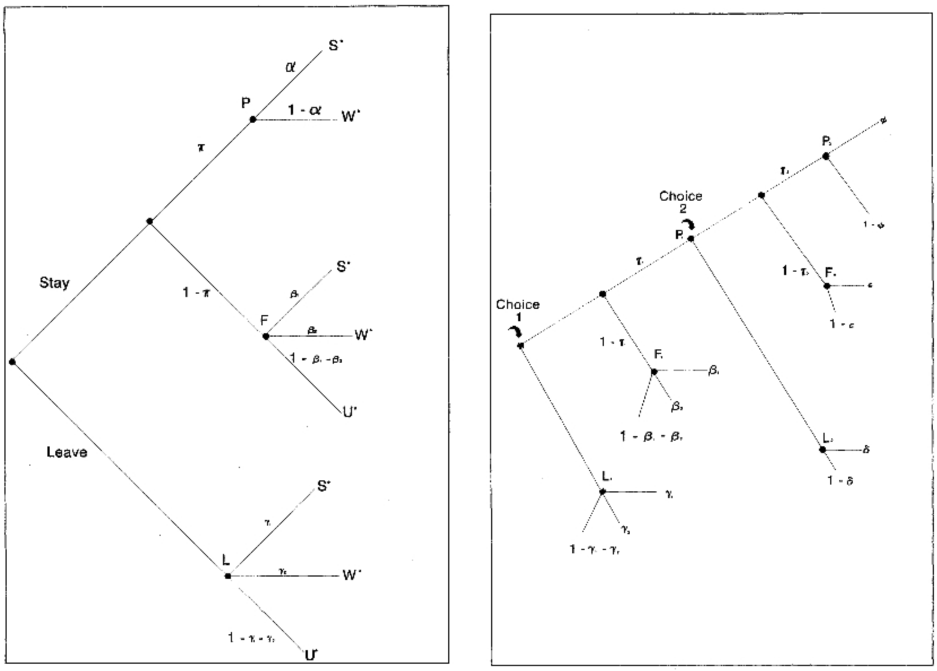
– while the corresponding probability of working-class children is given by

$$p_{iw} = \frac{\pi_i + (1-\pi_i)(\beta_1+\beta_2)}{\pi_i + (1-\pi_i)(\beta_1+\beta_2) + (\gamma_1+\gamma_2)}. \tag{2}$$

After a series of linear transformations, Breen and Goldthorpe (1997) can show that $p_{is} > p_{iw}$ for any value of π less than one (p. 284) – meaning that children from middle-class (i.e. from service and working class) will be more willing to continue a high level of education than to leave.

Now the authors impose two constraints: i) differences in both ability a_i and expectations of success π_i, and ii) differences in resources. With regard to i), following the assumption of primary effects of social inequality, the mean level of ability is higher in service class than in working class. If students' knowledge about their ability can be supposed to have a positive impact on their subjective probability assessments ($\pi_i = g(a_i)$), the average expectations of success π will be higher among service-class students (Breen and Goldthorpe, 1997, p. 285f.).

With regard to ii), if education can only be continued if the available resources r_i of each family exceed the costs of education c; and if service-class families dispose of greater



(a) Single decision tree. (b) Multiple decision tree

Figure 1: Single vs. multiple decision tree. Source: (Breen and Goldthorpe, 1997, p. 280, 288).

resources than working-class families, the share of students for which the condition $r_i > c$ holds is higher among service-class families than among working-class families.

However, (Breen and Goldthorpe, 1997) stress that the the theoretical model that is sketched by the single decision tree is a far too simplifying assumption. Empirically, a multiple decision tree would be the more adequate representation. Here, each decision about whether or not to continue onto educational level n may be made in the light of possible entry to educational level $n + 1$ (e.g. entry to university).

The multiple decision tree in figure 1 b pictures this situation for the case of two transition choices. Now there are five possible educational outcomes: immediately leaving from the lower level (L_1), staying at the lower level but failing the examination (F_1), the corresponding outcomes at the higher level (L_2 and F_2), and passing the higher-level examination (P_2). Notably, the higher-level outcomes are only open to those who pass the examination at the end of the lower level. Thus, both the likelihood-of-success assumptions and the utility assessments are conditional on the fact that there was no former dropout.[1]

[1] While Breen and Goldthorpe (1997) stress the "forward dependency" of transition decisions in the light of subsequent transition decisions, it shall be noted that the multiple decision tree also implies a kind of 'backward' dependency – meaning that each further transition is conditional on the antecedent transition decision(s) (also see Breen and Jonsson, 2000).

Esser's' Extension of the Breen-Goldthorpe Model Esser (1999) uses a SEU model to explain the mechanisms of parental educational choices at the end of primary school education. The expected utility EU for the alternatives at hand, to continue onto lower secondary school (A_n) or to continue onto intermediate or upper secondary school tracks (A_b) will be as follows:

$$EU(A_n) = P_{sd}(-SD) \tag{3}$$

$$EU(A_b) = P_{ep}B + (1 - P_{ep})P_{sd}(-SD) - C \tag{4}$$

with SD as the expected amount of status decline and with P_{sd} as its impact on parental decisions; B as the benefit of higher education (e.g. in terms of labour market prospects); P_{ep} as the subjective probability of successfully completing the chosen school track; and C being the expected costs of education (also see Becker, 2003; Pietsch and Stubbe, 2007). Esser (1999) can show that $EU(A_b) > EU(A_n)$ if $B + P_{sd}SD > C/P_{ep}$, while the term $B + P_{sd}SD$ can be denoted as the *educational motivation* and the term C/P_{ep} as the *investment risk*. Thus, a higher level of education will be aspired if the educational motivation to continue somehow exceeds the underlying investment risk.

Both the Breen-Goldthorpe- and the Esser model have been tested variously (Jonsson, 1999; Breen and Jonsson, 2000; Becker, 2003; Stocké, 2007; Schneider, 2008). However, as regards methods, Becker's (2003) operationalization controlling for selection bias via "Heckit" correction (Heckman, 1979) "(...) provides, at present, the best available test of the B[reen-]G[oldthorpe]-model" (Stocké, 2007, 508). In this model, first the impact of parental social class on each of the indicators B, $-SD$, p_{sd}, p_{ep} and C is used to correct for sample selection bias in their explanation of the choice of upper secondary school. Second, these effects are again used to control for selection bias in the explanation of the transition to particular school tracks (see section 3.4 for a more formal description of the Heckit correction).

Becker (2003) justifies his three-step method by the endogeneity of the causal structure. However, the next subsection will provide arguments that there is another endogeneity that has been neglected yet but is worthwhile to consider: the impact of teachers' expectations on the students' probability of successfully completing the chosen school track, p_{ep}.

2.2 Pygmalion in the Classroom

The idea of a *self-fulfilling prophecy* was firstly established by Robert Merton (1948). In this nowadays classical paper he showed how prejudices towards out-groups (e.g. African Americans) or specific attitudes about a certain situation (e.g. the rumor of a bank's illiquidity) might become true simply as a consequence of the former judgments: "The prophecy of collapse led to its own fulfillment" (Merton, 1948, p. 195). Following the well-known study of Rosenthal and Jacobson (1968), the effect of misled teacher expectations on student's future school achievement has been labeled as the 'Pygmalion Effect'.[2] The idea behind the metaphor is that too high or too low teacher expectations

[2]According to the Greek myth as it is narrated by Ovid (Metamorphoses, X), the Cypriot sculptor Pygmalion carved a woman out of ivory. This statue was so beautiful that he fell in love with it.

will have an impact on teacher-student interactions, which, in turn, might influence the students to adopt their motivations and aspirations according to their teachers' expectations. In the words of Merton, originally misled teacher expectations will have led to their own fulfillment.

The classical Pygmalion In the original study, Rosenthal and Jacobson (1968) administered a nonverbal intelligence test to elementary school children. However, they did not tell the teachers that this was an intelligence test, but that it was a new tool to identify 'late bloomers', i.e. children who were likely to show a sudden and dramatic intellectual spurt over the upcoming school year. But although the 'late bloomers' were actually selected at random, Rosenthal and Jacobson (1968) could show that in an IQ test which was administered one year later they gained significantly better test scores than the control students. Thus, the false expectations of the teachers (who had been informed about the artificially created group of late bloomers) had become true.[3] Whereas many social psychologists took Pygmalion as a confirmation of their thesis that social reality is mainly created by one's own expectations, educational psychologists were much more skeptical with regard to Pygmalion's methodological prerequisites and the possibility of alternative explanations which, according to them, Rosenthal and Jacobson (1968) have not sufficiently controlled for (Jussim and Harber, 2005, 139).[4]

Trying to refute his critics, Rosenthal became one of the pioneers in meta-analyses. His and Rubin's (Rosenthal and Rubin, 1978) meta-analysis of the first 345 studies from various research categories (reaction time, inkblot tests, animal learning, laboratory interviews, psychophysical judgments, learning and ability, person perception, and everyday life situations) concluded that self-fulfilling prophecies do exist and show effect sizes between d = .14 up to d = 1.73 and r = .07 up to r = .65 (Rosenthal and Rubin, 1978, table 1). A second meta-analysis based on a more narrowly defined set of 'Pygmalion' studies could examine that the effect of teachers' expectations on students' IQ scores was .16 by average (Smith, 1980). Raudenbush (1984) found an effect size of .11 by average and could additionally reveal that the effect of teachers' expectations at t_0 on later IQ scores at t_1 highly depends on how long the teachers are already teaching in the particular class.[5]

Due to his caress, the statue finally gets alive, they marry and have a son.

[3] While social psychology differentiates between the *Pygmalion* effect of self-fulfilling *over*-estimations and the *Golem* effect of self-fulfilling *under*-estimations, we use the more common term of Pygmalion to capture both types of self-fulfilling prophecies.

[4] Critics remarked that both groups of children – late bloomers and controls – showed IQ gains over the next year. The differences between the gains of the two groups (four percentage points) are significant, but less 'dramatic' than the gross IQ gain of 12 percent of the experimental group students would suggest. For this and other critiques with regard to the original Rosenthal and Jacobson (1968) study see Thorndike (1968); Jensen (1969); Snow (1969); Elashoff and Snow (1971); Wineburg (1987); Roth (1995) and Jussim and Harber (2005).

[5] A duration of less than 5 weeks can yield to an effect size of up to .55, whereas a duration of 24 weeks led to an effect size of -.13 (Raudenbush, 1984, 91). Thus, the longer a teacher is teaching in a particular class, the better he knows his or her students and the smaller are the consequences of possible misjudgments.

Although critics like Wineburg (1987) refused to accept an impact of teachers' expectations on students' intelligence scores, Raudenbush (1994) re-analyzed the 18 experiments of his earlier study (Raudenbush, 1984) based on random effect models and now found an effect size even of r=.20.

Need for moderators Given these results, one evident weakness of 'Pygmalion' regardless of its operationalization lies in an insufficient control of both student and teacher background variables as moderators.[6] In particular, more research is clearly needed with regard to students' social backgrounds as moderators (Jussim and Harber, 2005). Concretely, there are only three studies who explicitly tested for these effects: First, Madon et al. (1997) found that self-fulfilling prophecies are stronger among students who had a 'prior history of low-achievement' which was operationalized as their standardized results of a test which had been administered previously to the actual experiment. Although their operationalization of self-fulfilling prophecies as teachers' over- and under-estimations – in terms of the residuals of a regression of three different teacher perception variables (related to students' performance, talent and effort) on a set of student background variables – appears to be promising, for their purpose of identifying moderator variables it surely suffers from a serious methodological weakness.[7]Second, Jussim et al. (1996) found evidence that self-fulfilling prophecies are moderated by both social class and ethnicity variables. In their study, the standardized relationship between teachers' perceptions and students' future test scores was about .25 for students with lower educated parents and .03 for students with higher educated parents. Similar differences could be detected between white students and African-American students in terms of a standardized effect size of .14 and .37, respectively. Third, Madon et al. (1998) found that teachers' perceptions about students' performance and talent (but not about their ability) correlate bivariately with students' social class (operationalized as an index of parental education and parental income). However, these bivariate associations diminish when additional predictors like students' school grades, intelligence test scores and their motivation are introduced in multivariate analyses. Hence, the biggest part of the differences that teachers find between social groups closely correspond to actual differences in prior grades and achievement tests.

Implications What does this overall mixed evidence suggest? First, the phenomenon of a self-fulfilling prophecy is hard to identify analytically. As we saw, not only experimental-group students achieved a gain in their IQ test scores but also control-group students

[6]Among the few exceptions of empirical studies which take moderator effects into account, the meta-analysis by Raudenbush (1984) which found that the effect size of self-fulfilling prophecies varies with teachers' *duration in class* has already been mentioned. Moreover, in the same study Raudenbush (1984) found that the effect also varies by *grade level*. And finally, self-fulfilling prophecies appear to be weaker in more 'differential' *teacher treatment contexts* (Brattesani et al., 1984).

[7]Concretely, among the set of background variables that was used to identify teachers' over- and under-estimations, we can find students' fifth-grade math test scores – which were also used to identify low and high achievers (Madon et al., 1997, 798). Therefore, it is not surprising that the authors find a variation in the effect size of self-fulfilling prophecies based on this variable.

(Rosenthal and Jacobson, 1968). Second, we can note that one solution might be to compute a 'net' effect of self-fulfilling prophecies in the way of Madon et al. (1997). Although those strategies evidently are not without pitfalls, they might be helpful in separating self-fulfilling prophecy effects from other intervening mechanisms. Third, and most important, we saw that self-fulfilling prophecy research lacks of a sufficient consideration of student background variables. Exactly this is the adequate point to bring the SEU-IEO framework back in: Just as self-fulfilling prophecy research needs the consideration of student background variables, the SEU-IEO framework lacks the consideration of exactly that endogeneity of the probability of educational success which is the main point of all *Pygmalion* studies. The task of the next section will be to integrate the main idea of a 'net' effect of a self-fulfilling prophecy into the SEU-IEO framework.

2.3 Development of an SEU model of self-fulfilling prophecies

Given the utility relations of the conventional SEU-IEO model as it has been outlined in section 2.1, educational decisions would be a direct function of net utility. However, this seems to be only half of the truth for it would neglect the idea of a self-fulfilling prophecy in the classroom. In line with the main idea of *Pygmalion*, claiming that a teacher's expectations may have a distinct effect on later school achievement implies that the 'real' transition rates are not only a result of 'subjective' parental utility comparisons, but also of 'objective' interactions in the classroom: "A shortcoming of the standard economic approach to decision making is that it ignores the endogeneity of preferences - that students' preferences are socially constructed through interaction with peers and other significant persons" (Lauen, 2007, 183). The consequence of admitting an endogeneity of preferences in the classroom is to assume also an endogeneity of p_{ep}, i.e. of the subjective probability of successfully completing the chosen school track. Recall that following Breen and Goldthorpe (1997), the subjective probability of educational success depends on students' objective school performance (in their notation $\pi_i = g(a_i)$). In accordance with Esser's (1999) notation, we could write

$$P_{ep} = f(AP), \tag{5}$$

while AP denotes students' academic performance. Claiming that teacher expectations in terms of a 'net' effect of self-fulfilling prophecies (Madon et al., 1997) at time $t-1$, TE_{t-1}, may influence students' academic outcomes at a later time t reduces to

$$AP_t = g(TE_{t-1}). \tag{6}$$

For P_{ep_t} thus holds

$$P_{ep_t} = f(g(TE_{t-1})) \tag{7}$$

– meaning that parental subjective probability assumptions are a function of students' objective school performance which is, in terms of a self-fulfilling prophecy, dependent on the earlier teacher expectations. If we apply this idea on Esser's (1999) formal model and simply abbreviate the relation in (6) by SFP to indicate that it captures the idea of a self-fulfilling prophecy, we can write it as follows:

$$EU(A_b) > EU(A_n)$$

if

$$B + P_{sd}SD > C/P_{ep}$$

with

$$P_{ep} = f(SFP). \tag{8}$$

Now we can infer that self-fulfilling prophecies will also affect students' later transition decisions via their impact on the subjective expected performance assessments.[8]

2.4 Hypotheses

After these theoretical considerations our main hypothesis is easily outlined: We postulate that teachers' expectations - once they are operationalized in terms of a self-fulfilling prophecy - have distinct effects on students' educational success. By 'distinct effects' we mean that they will have a significant impact besides the convenient theoretical concepts of the SEU-IEO model (Esser, 1999). While the latter has usually been applied to the transition decisions from primary to secondary education[9], one central claim of the initial Breen-Goldthorpe model (Breen and Goldthorpe, 1997) is that secondary effects of social inequality do not only affect the actual transition decisions but also the decision for or against *continuing* the chosen school track.[10] Consequently, in the first step of our analyses we will use the SEU-IEO model to predict the probability of 10^{th} class students to achieve a German high school degree ('Abitur'). In a second step, we will also model their transition probability to tertiary education in terms of starting academic studies.

Therefore, we will test two sets of hypotheses consisting each of 8 single statements, $H_{1a} - H_{8a}$ and $H_{1b} - H_{8b}$. The first five hypotheses of each set test the standard SEU-IEO model for students' probability to achieve a German high school degree ($H_{1a} - H_{5a}$) and to start academic studies ($H_{1b} - H_{5b}$), respectively. H_{6a}, H_{7a}, H_{6b} and H_{7b} address the combined SEU terms which Esser (1999) calls *educational motivation* and *investment risk*. The last hypotheses of each set, H_{8a} and H_{8b}, extend the SEU-IEO model by the idea of a distinct effect of teacher expectations in terms of a self-fulfilling prophecy:

H_{1a} : Students' probability of achieving a high school degree increases with the benefit of higher education, B.

[8]One question that might arise at this point is why self-fulfilling prophecies should have a distinct impact on students' academic outcome apart from an objective measure of their academic performance at $t-1$, AP_{t-1}. The answer to this point is given in section 3.3, when the idea of a net effect of self-fulfilling prophecies will be operationalized as the residuals of a regression of teacher evaluations on a performative component, and a motivational component, respectively (see Madon et al., 1997).

[9]Exceptions are Jonsson (1999); Need and De Jong (2001); Becker and Hecken (2008, 2009), and, without a direct operationalization of the SEU indicators, Hillmert and Jacob (2010).

[10]See Schneider (2008) for an empirical test of this hypothesis.

H_{2a} : Students' probability of achieving a high school degree increases with the value of status decline, $-SD$.
H_{3a} : Students' probability of achieving a high school degree increases with the expected probability of status decline, P_{sd}.
H_{4a} : Students' probability of achieving a high school degree increases with the expected educational performance, P_{ep}.
H_{5a} : Students' probability of achieving a high school degree decreases with the expected costs of education, C.
H_{6a} : Students' probability of achieving a high school degree increases with their educational motivation, $B + p_{sd} * SD$.
H_{7a} : Students' probability of achieving a high school degree decreases with their subjective investment risk, C/p_{ep}.
H_{8a} : Students' probability of achieving a high school degree increases with (positive) self-fulfilling prophecies, SFP.

H_{1b} : Students' probability of beginning academic studies increases with the benefit of higher education, B.
H_{2b} : Students' probability of beginning academic studies increases with the value of status decline, $-SD$.
H_{3b} : Students' probability of beginning academic studies increases with the expected probability of status decline, P_{sd}.
H_{4b} : Students' probability of beginning academic studies increases with the expected educational performance, P_{ep}.
H_{5b} : Students' probability of beginning academic studies decreases with the expected costs of education, C.
H_{6b} : Students' probability of beginning academic studies increases with their educational motivation, $B + p_{sd} * SD$.
H_{7b} : Students' probability of beginning academic studies decreases with their subjective investment risk, C/p_{ep}.
H_{8b} : Students' probability of beginning academic studies increases with (positive) self-fulfilling prophecies, SFP.

It should become clear at this point that our aim is not to test *Pygmalion* explicitly. Due to data restrictions, we are not able to test for a direct impact of teacher expectations on students' future school performance. However, the considerations from section 2.3 suggest that given an adequate operationalization of self-fulfilling prophecies, in the long run we can expect an indirect effect of teachers' expectations on both students' educational success and transition decisions via their (unobserved) subjective expectations of educational success at a later point in time. Exactly this is what H_{8a} and H_{8b} are trying to capture.[11]

[11] In this context, one could also refer to the distinction between *substantive* and *empirical* statistical models (Cox, 1990), or between scientific models presented in statistical form and statistical models *per se* (Rogosa, 1987; Sø rensen, 1998). The point is that the former "are intended to represent real processes that have causal force (whether or not directly observable)" while the latter "are

3 Operationalization

3.1 Data

All analyses will be based on a German panel dataset which is known as the 'Kölner Gymnasiasten-Panel' (engl. 'Cologne Highschool Panel', in the following abbreviated as *CHiSP*). The CHiSP consists of an initial (student-level) survey from 1969 (Gesis-No.: ZA0600) with N=3385 10^{th}-grade Gymnasium[12]students in North Rhine-Westphalia with two re-surveys in 1985 (Gesis-No.: ZA1441; N= 1987) and 1996/97 (Gesis-No.: ZA4228; N=1596). In the initial survey, students were asked about issues like their performance, interests and plans in school and about their social origin and their relationship to their parents. Simultaneously with the initial survey, the students took part in an Intelligence Structure Test (IST) containing four sub-scales as developed by Amthauer (1957). At the same time, also the students' teachers (Gesis-No.: ZA0640; N=1701) and their parents (Gesis-No.: ZA0639; N=2646) have been surveyed. The main items of the parent questionnaire were about their social background, their style of raising children and their aspirations for their children. Amongst others, teachers were asked about a couple of evaluative and other pedagogic issues. In an investigation of the Central Archive for Empirical Research in Cologne (today known as Gesis - Leibniz Institute for the Social Sciences) the 10^{th} class and Abitur grades (if passed) could be examined and were merged with the data. In the two re-surveys, the former students gave detailed information about their educational and occupational careers until the age of 43. We chose these admittedly older data because to the best of our knowledge, it is the only available longitudinal dataset that contains an appropriate measure of teachers' expectations that can be used to construct over- and underestimations in order to operationalize self-fulfilling prophecies adequately. This indicator will be described in the next but one paragraph.

3.2 Variables

Dependent Variable In the hypotheses section we identified two dependent variables. The first dependent variable is defined by the fact if the students have achieved a high school degree (Abitur) or not. While the CHiSP also includes information about whether the former students have *ever* achieved Abitur in their later life, we will focus on those students only who achieved Abitur during the regular schooling time. This appears to be logically consistent since secondary effects of social inequality can also be understood as a decision for vs. against continuing higher education (Breen and Goldthorpe, 1997; Schneider, 2008). Hence, we want to focus only on students who passed Abitur on the first try (event=1) use all observations who did not achieve Abitur within 3 years after

those which sociologists normally use and are concerned with relations among variables that may be determined through techniques of rather general applicability" (Goldthorpe, 2001, p. 14).

[12]For more detailed descriptions of the German educational system see Jürges and Schneider (2006); Pietsch and Stubbe (2007); Schneider (2008).

the 10^{th} class survey in 1969 as a reference (=0).[13] The second dependent variable is given by the fact if the former students have ever started academic studies. Since our analyses will be based on panel data, we have to take into account that from a theoretical point of view it would be possible that the former students could start academic studies at any time – including data points that are later than the last survey of the CHiSP (i.e. 1997). This problem will be solved empirically in section 4.1.

Independent Variables The *expected benefit* of education, B is operationalized by students' appraisement if Abitur would be necessary for them to reach their aim in life. Students could reply 1 'yes, necessary'; 2 'useful, but not necessary'; and 3 'not important'. We dichotomized this variable into the two categories 0 'not important' and 1 'useful or necessary'. The *value of status decline*, $-SD$, is measured by parents' disappointment if child would not pass Abitur. The categories of this variables are 1 'not much'; 2 'little'; 3 'very disappointed'; 4 'would be the worst'. We dichotomized this variable as follows: 0 'not much / little'; 1 'very disappointed / would be the worst'. We operationalize the *expected status decline*, p_{sd} by parents' assessments about the importance of good Abitur grades for students' later occupational success. The original categories of this variable (1 'little'; 2 'not that much'; 3 'big'; 4 'very big') were dichotomized into 0 'little / not much' and 1 'big / very big'. Students' *subjective educational performance* p_{ep} is measured by a probability assumption of the parents whether their children are able to complete the chosen school track. The original variable (1 'definitely'; 2 'probably'; 3 'don't know'; 4 'probably not') is recoded as follows: 0 'probably not/don't know'; 1 'probably/definitely'. The *expected costs* of education, C, are operationalized by parents' assessment if they had to make financial sacrifices in order to offer higher education to their children. Again, the original categories of the variable (1 'no', 2 'little' and 3 'yes' are recoded into a dummy variable: 0 'no/little'; 1 'yes'.

Self-fulfilling prophecies, SFP, should adequately be operationalized based on teachers' expectations. In the CHiSP the latter are measured by a specific form of teachers' evaluations: Teachers were asked to evaluate by a dichotomous decision which students they suppose to be able for academic studies and which of them not. Since the question was put open, teachers could mention students as being able, being not able, or not at all.

This data structure causes two problems. First, each student could be evaluated by more than one teacher, and each teacher could evaluate more than one student. An analysis of the intra-class correlations (ICC) revealed a considerable variance of multiple evaluations for each student (not shown). Second, the openness of the question is not without problems because it has to be clarified whether the 'missing' category really can be treated technically as a missing value or if we would loose substantive information when proceeding on this assumption.

[13] Since the zero point of counting has been backdated to January 1967 and we do not want to exclude students who had to participate in makeup exams, we set the cut-off value to 80 months beginning from the starting point.

To overcome the first problem, our analysis here will focus on teacher evaluations only of class teachers. We expect that the intra-individual variance of teachers' evaluations partially depends on the quality of teacher-student relationships. We assume that class teachers have a more intense relationship to and a better knowledge of their students than 'ordinary' teachers. Thus, regarding only class teachers' evaluations will both simplify the data structure and overcome the problem of variance.[14] To overcome the second problem, as a preliminary analysis we have estimated two logistic regressions of the chance of getting a positive evaluation vs. getting a negative one or none at all on students' intelligence, average grade, social background, motivation and sex (not shown). We found that for the analysis of the chance of getting a positive evaluation vs. not getting one at all, the effect sizes of all independent variables are in the same direction, but notably lower than for the analysis of the chance of getting a positive evaluation vs. getting a negative one. Thus, we can conclude that students who are not mentioned at all rank *lower* in teachers' perceptions than students with a good teacher evaluation but *higher* than students with a bad teacher evaluation. However, in these analyses we will look at the unambiguous values of this variable in terms of the opposition of positive vs. negative teacher evaluations.

Based on this dichotomy SFP is measured as follows: Teachers' evaluations are regressed on two sets of students' backgrounds: an ability component and a motivational component. The ability component consists of students' scores in the Intelligence Structure Test (Amthauer, 1957) and their average grade (both z-transformed). The motivational component consists of students' subjective assessments of i) their homework effort, ii) their relative school performance, and iii) their self-confidence (all 11-point Likert scaled). Teachers' evaluations are subsequently regressed on these two sets of student backgrounds, resulting in three different logistic regression models: one for each set, and a 'full' model with all predictors. The models read as follows:

$$logit(TE) = \beta_0 + \beta_1 intell + \beta_2 av.grade \tag{9}$$

$$logit(TE) = \beta_0 + \beta_3 homew.eff + \beta_4 subj.rank + \beta_5 self.conf \tag{10}$$

$$\begin{aligned} logit(TE) = \beta_0 &+ \beta_1 intell + \beta_2 av.grade \\ &+ \beta_3 homew.eff + \beta_4 subj.rank + \beta_5 self.conf, \end{aligned} \tag{11}$$

where (9) denotes the performance model, (10) the motivation model, and (11) the full model. The residuals of (9) to (11) are stored and will be used as predictors for students' propensity to achieve Abitur and to start academic studies, respectively (see Madon et al., 1997 for a similar operationalization of self-fulfilling prophecies). The residuals r_i of logistic regressions are defined as $r_i = y_i - logit^{-1}(X_i\beta)$ (Gelman and Hill, 2007, p. 97), where in our case y_i is the observed teacher evaluation and $logit^{-1}(X_i\beta)$ is the value of each teacher's evaluation that is predicted by equations (9) to (11). In this design,

[14] As regards social mechanisms, we further expect that class teachers' evaluations might very well be an approximation of teachers' evaluations in general: There is good reason to presume a notable amount of communication *between* teachers, e.g. in the teachers' lounge, and especially class teachers could be *agenda setters* in terms of shaping other teachers' expectations.

positive residuals indicate relative over-estimations and negative residuals relative under-estimations compared to the respective set of predictors in the logit models. For later analyses we will dichotomize each residual whether it takes positive or negative values. By this procedure, it is possible to separate a 'net' effect of self-fulfilling prophecies from a varying set of background variables (Madon et al., 1997).[15]

As regards the distinction between *primary* and *secondary* effects of social inequality, in both the performance model and the full model, differences in the distribution of students' teacher evaluations which are due to primary factors like their intelligence are explicitly ruled out. According to the assumptions in the theoretical model, the effect of the residuals from these models on students' actual transition propensities should only exist due to a mechanism of secondary effects in terms of different subjective expected transition probabilities – which are in turn assumed to be the consequence of different teacher treatments.

3.3 Covariates

To keep a check on the unobserved heterogeneity of our predictors (also see section 4.3), our analyses control for parental social class and educational attainment. *Social class* is measured by the occupational prestige (Treiman scores) of the head of household - while the latter is based on a variable that takes the highest value of occupational prestige from either mother or father.

The *parental educational attainment* was measured by 13 categories reaching from lower secondary school without an apprenticeship up to a university degree. We categorized this variable into 1 'lower education'; 2 'middle education'; 3 'higher education' and 4 'degree'.

3.4 Models

In our models we will mainly follow the operationalization that has been provided by Becker (2003). First all predictors will be regressed on parental social class via Probit estimations. The estimates will be stored as Inverse Mill's Ratios (IMRs) and will be introduced in the second-step logit estimation of students' probability of passing Abitur to control for panel mortality (Heckman, 1979). This will be repeated for students' propensity to start academic studies.

The general assumption of this statistical technique is that in many social situations the outcome of primary interest y_i not only depends on a vector of covariates β but

[15] An issue that could be objected against this strategy concerns the possibility of *private information*. More specifically, besides the variables in the three models, teachers could ground their decisions on two different types of unobservables: a component that is known to the teacher when she makes an evaluation decision but not to the analyst, and a component that might not even be known to the teacher herself (Cunha et al., 2005; Cunha and Heckman, 2007). While it can be argued that the latter case would be in line with the general idea of a self-fulfilling prophecy (although the particular mechanism behind it would remain un-revealed), we tackle the implications of the former scenario in our robustness analyses in section 4.3.

also on a variable z_i that determines whether individual i will ever entry in the social situation or not. Thus, the crucial assumption is that we will only observe y_i if $z_i > 0$, and therefore we first have to find the determinants of z_i (on which the latter should be regressed) before we can say anything about the relationship between β and y_i. In more formal terms, we can distinguish between a *selection equation*,

$$z_i = \gamma' w_i + u_i,$$

and the equation of primary interest,

$$y_i = \beta' x_i + \epsilon_i.$$

Since, as mentioned, y_i is only observed if $z_i > 0$, the error terms of both equations, u_i and ϵ_i, share the correlation ρ. The consequence of this error correlation is that conventional OLS regression that only considers equation 2 produces inconsistent and inefficient estimates (cf. Greene, 2003; Wooldridge, 2006).[16] This selection problem can be solved by a two step estimation of both equations: First the selection equation is estimated via Probit regression:

$$Prob(z_i = 1|\mathbf{w}_i) = \Phi(\mathbf{w}_i'\gamma)$$

and

$$Prob(z_i = 0|\mathbf{w}_i) = 1 - \Phi(\mathbf{w}_i'\gamma).$$

Then the estimates of γ' are stored as *Inverse Mill's Ratio* λ_I which are computed as follows:

$$\lambda_i = \phi(\mathbf{w}_i'\gamma)/\Phi(\mathbf{w}_i'\gamma).$$

In a second step, λ_i is included as a covariate in the equation of primary interest. This two-step procedure is intended to yield a more precise estimate of β_i because by controlling for λ as a metric instrumental variable for the exogenous determinants of the selection equation, also the problematic error correlation ρ is canceled out.

In our case, there is evidence to assume two different sources of sample selection in the data. First, the distribution of our independent variables is expected to vary by parental social strata (Becker, 2000, 2003): Social backgrounds might affect both the definition and evaluation of the social situation, and thus also the decision for or against a higher track of education. Therefore, by regressing all SEU predictors on parental social class (selection equation) and including the IMRs of these estimates in the equation of interest, we are able to control for the causal impact of class-specific resources, conditions and constrains:

> From the methodological point of view, the following aspects are considered separately: (1) the unobserved heterogeneity based on the interrelation between social class and social action; (2) the social selectivity of resources, educational preferences, and educational performance among social classes; (3) the social selectivity

[16] In this case, OLS estimates are *inconsistent* due to the omitted variable w_i and *inefficient* due to the heteroscedasticity in terms of ρ.

of the evaluation of the costs and benefits of continued education; and (4) the problem of causal inference in the decision problem (Becker, 2003, p. 15).[17]

With regard to our self-fulfilling prophecy residuals, recall that one central shortcoming of the empirical literature of *Pygmalion* affects an insufficient consideration of students' social backgrounds. By applying the same Heckit model on the dichotomized SFP residuals, it is possible to control for the variance of self-fulfilling prophecies by parental social class.

Second, evidently the propensities of the former students to start academic studies strongly depend on the fact whether they successfully graduated from Gymnasium or not. Although for some particular subjects like music or art a special qualifying examination can substitute a high school degree, in most cases transitions to university can only be observed if Abitur has been passed successfully. Thus, the theoretical problem of conditional transition rates (Breen and Goldthorpe, 1997) should also be solved as a methodological problem of selection bias.

Concretely, we will estimate the following models: First, as mentioned, parents' expected benefit of higher education B, their subjective value of status decline $-SD$, the expected status decline p_{sd}, students' probability of successfully completing the chosen school track p_{ep}, the expected costs of education C, and our dichotomized indicator for self-fulfilling prophecies SFP will be regressed on parental occupational prestige in bivariate probit equations to compensate for social selection bias in the distribution of our predictors. The results of these probits will be stored as Inverse Mill's Ratios λ_{ij}. By subscripts i, j we address that each individual i will get an own Inverse Mill's Ratio for each selection equation j. In a second step, each λ_{ij} is introduced in our first equation of primary interest, which predicts the individual probabilities to pass Abitur:

$$logit_{ABI} = p_{ep}B + (1 - p_{ep})(-SD) - C + SFP + \lambda_{ij}. \tag{12}$$

Second, we will re-estimate (12) as a probit equation and equally store the resulting estimates as IMRs in order to control for sample selection with regard to the transition rates to university. In addition to (12), this model also includes *direct* controls for parental social class and educational backgrounds:

$$logit_{UNI} = p_{ep}B + (1 - p_{ep})(-SD) - C + SFP + \lambda_{ij} + class + educ. \tag{13}$$

This procedure is summarized graphically in figure 2.

[17] In our case, this problem might be even stronger since our data only contain records of Gymnasium students.

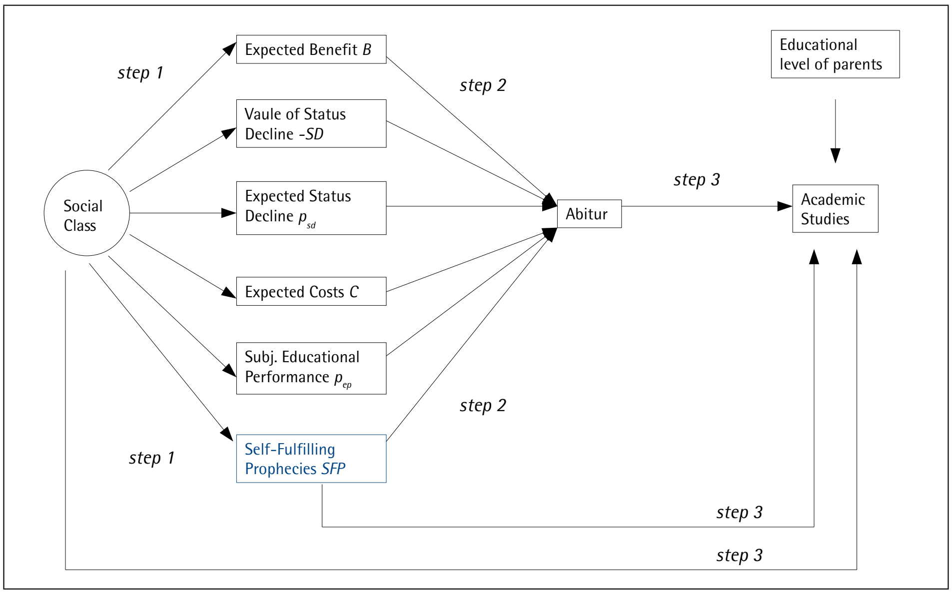
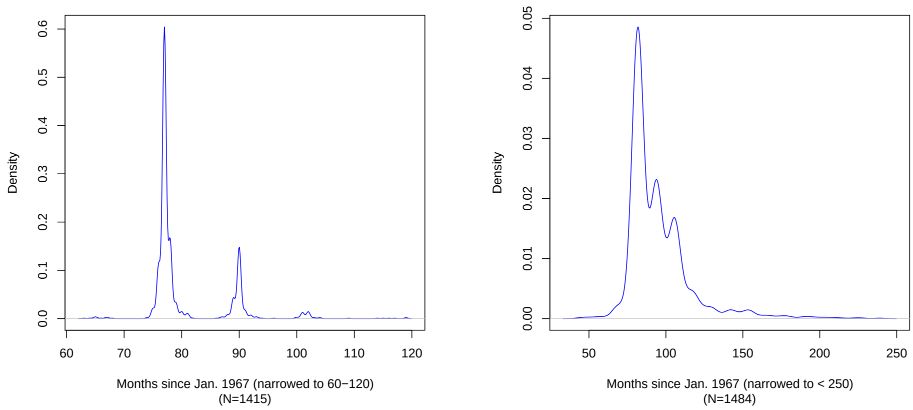


Figure 2: A modified model of Inequality in Educational Opportunities. Own Contribution. Extension of Becker (2003, p. 7).

4 Results

4.1 Distribution of Variables

Dependent Variables In figure 3a the distribution of the time span until students passed Abitur is displayed. Recall that the zero point of counting has been backdated to January 1967. We can see that the distribution of passing Abitur over time corresponds to our cut-off value of 80 months. Most of the students passed Abitur on the first try, a quite small amount on the second try, and even less on the third try. Figure 3b captures the distribution of the time span until the former students have started academic studies. Most of the students started academic studies immediately after having passed Abitur, and some smaller amounts with a delay of one to two years. After 106 months beginning from the starting point – which is equal to October two years after high school graduation – the amount of students who begin academic studies tremendously drops down. Thus, we choose this value as the cut-off for dichotomization of our second dependent variable.



(a) Distribution of educational success over time (b) Distribution of university transitions over time

Figure 3: Distribution of educational success and university transitions over time

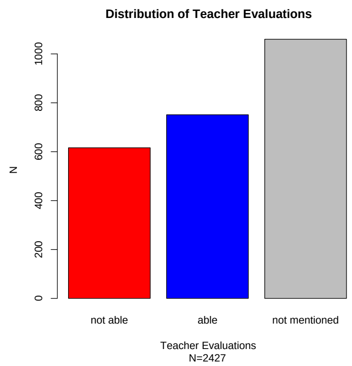
Main Independent Variable: Teacher Evaluations Now we present the distribution of teachers' evaluations both numerically (figure 4a) and graphically (figure 4b). It can be noted that the amount of students who received a positive teacher evaluation (30.9%) is higher than the amount of students who received a negative one (25.4%) – but evidently most students did not receive any teacher evaluation at all (43.7%). As mentioned in section 3.2, for the following operationalization of self-fulfilling prophecies we will only focus on positive vs. negative teacher evaluations.

Residuals of Over- and Under-Estimations Next we present the results of our logit equations (9) to (11) that we use to extract the 'net' effects of self-fulfilling prophecies.

	teacher evaluation			
	not able	*able*	*not mentioned / missing*	*Row Total*
n	616	751	1060	2427
% students	25.4	30.9	43.7	100.0
% mentioned	45.1	54.9	-	-

N(total) = 3385.

(a) Numerical distribution of teacher evaluations



(b) Graphical distribution of teacher evaluations

Figure 4: Distribution of teacher evaluations: ability for academic studies

Model 1 contains the performance model, model 2 the motivation model, and model 3 the full model with all predictors from both models 1 and 2 (see table 1).[18]As regards the performance model (model 1), we can note that both students' intelligence and their school grades significantly predict teacher evaluations. The R^2 of this model is remarkably high. However, except of the measure for students' relative school performance, in the motivation model, the z-values are much lower (self-confidence) or do not even reach statistical significance (homework effort). This also results in an R^2 not much more than half as high as for the performance model. Considering the predictors of both models together, in the full model, except students' relative school performance, only performance-model indicators remain significant – while the explained variance of the full model is only slightly higher than for the performance model. Thus, we conclude that students' performance is far more important for the teachers than their motivation.

As mentioned, we now store the residuals in order to use them as an operationalization of a 'net' effect of self-fulfilling prophecies. Figure 5 displays the distribution of the residuals from the three different logit models. Positive residuals indicate an over-estimation relative to the predictors of the logit models, negative residuals a relative under-estimation. In accordance with the predictive power of the performance model, the residuals in figure 5a mainly follow a normal distribution: Most students get a teacher evaluation that is quite in line with their intelligence and school grades – leading to a residuum of zero. If we compare this distribution with the one of the residuals from the motivation model, we can note that students' motivation hardly suits to predict teachers' evaluations: Two local maxima can be found at 0.5 and -0.5, respectively – indicating that based on this background variables, the teacher evaluations do not become better than simply by guessing. Finally, when we look at the distribution of the full-model residuals, we see that the curve gets slightly distorted, but still is very close to the normal distribution. As noted before, we dichotomized each residual for the

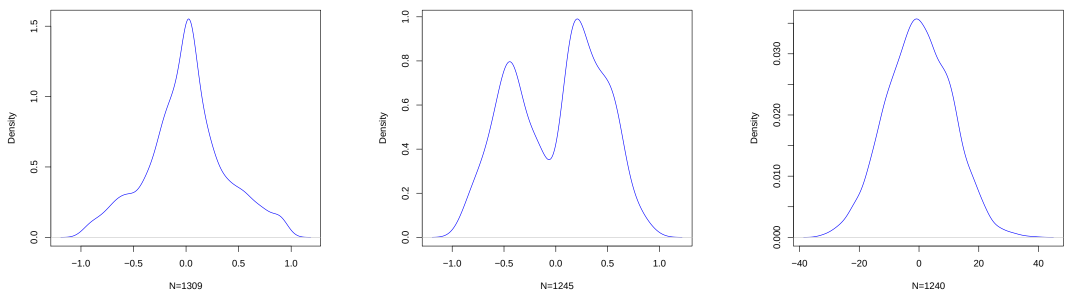
[18]For these and all subsequent models, missing values have been deleted listwise (for more details see Allison, 2002).

Table 1: Logistic Regression of Teacher Evaluations on Students' Performance and Motivation

	Performance Model e^{b*sd}/z-value	*Motivation Model* e^{b*sd}/z-value	*Full Model* e^{b*sd}/z-value
intelligence	1.76***		1.79***
	(7.00)		(6.90)
average grade	0.15***		0.20***
	(-16.79)		(-14.01)
homework effort		1.03	1.08
		(0.50)	(0.88)
relative school performance		2.74***	1.85***
		(11.11)	(5.65)
self-confidence		1.22*	1.10
		(2.46)	(1.01)
Nagelkerke's R^2	0.52	0.27	0.55
N	1309.00	1294.00	1287.00

All coefficients are standardized odds ratios. Z-values in parentheses. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

following analyses.

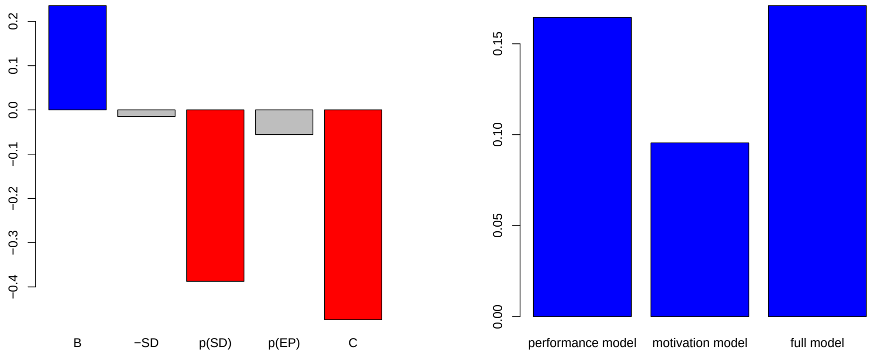


(a) Performance Model (b) Motivation Model (c) Full Model

Figure 5: Distribution of Residuals of Teacher Evaluations on students' performance and motivation, and the combination of both

Bivariate Probit Estimates In this section we briefly discuss the results from the bivariate probit regressions of our SEU predictors of primary interest, B, p_{ep}, SD, p_{sd}, and C on parental social class (see figure 6a). Blue bars indicate significantly positive coefficients, red bars significantly negative coefficients, and grey bars insignificant coefficients.

Among the SEU predictors, only the expected benefit B is positively predicted by parental social class – meaning that parents from higher social strata expect more ben-



(a) SEU Predictors

(b) SFP Residuals

Figure 6: Standardized Bivariate Probit Estimates of SEU Predictors and Self-Fulfilling Prophecy Residuals On Parental Social Class

efit from higher education. Not surprisingly, social class is negatively related to the subjective assessment of the costs of education. What might surprise, however, is that social class is also negatively related to parents' assessment of the probability of the impact of the expected status decline: While parents from lower social strata seem to be more concerned about the potential impact of a lacking Abitur on their children's' later life, parents from the higher strata appear to show much more trust that due to their higher resources they would be able to make straight paths for their children even in case of a failed final exam. With regard to $-SD$ and p_{ep}, no significant associations were found.

Figure 6b shows that all three types of residuals are positively predicted by parental social class. Hence, students from the higher social strata are more likely to be over-estimated by their teachers compared to their actual performance and motivation.

4.2 Multivariate Analyses

First we present the logistic regression estimates of students' probability of passing Abitur on both Esser's (1999) SEU predictors and our self-fulfilling prophecy residuals. We subsequently introduce our predictors and correction terms in our models, so that we will present the following models: *Model 1a* contains the convenient SEU-Predictors (Esser, 1999) B, $-SD$, p_{sd}, p_{ep} and C. In *model 2a* to *4a* we separately introduce the performance residuals, the motivation residuals, and the full-model residuals in order to model the impact of self-fulfilling prophecies. *Models 1b* to *4b* contain the same variables as models 1a - 4a but additionally correct for sample selection bias in terms of the Inverse Mill's Ratios that have been stored from the bivariate probit models from figure 6. *Models 1c* to *4c* repeat the same proceeding for the terms that Esser calls "educational motivation", $B + p_{sd} * SD$, and the "investment risk", C/p_{ep}. And *models 1d* to *4d* additionally control for potential selection bias in models 2c to 4c. Since the regression models with separate IMR variables seriously suffered from multicollinearity (the inter-correlation between the IMRs lies between an absolute value of .97 and .99), we summed up the IMR scores for all SEU predictors to one single IMR score.[19]

Second, in *models 5a* to *8d* we present the estimates of another series of logistic regressions of students' transitions to university. The setup of these models is the same as in models 1a to 4d, except that models 5b to 8b and 5d to 8d now include the Inverse Mill's Ratios for the estimates of a probit version of models 1b-4b and 1d-4d, respectively. Since the results for the self-fulfilling prophecy residual estimates do not substantively differ when the SEU interaction terms instead of the "main indicators" are introduced in the model, the tables with the interaction terms are not discussed in depth here but reported in the appendix (tables B and C).

Passing Abitur As table 2 shows, in the baseline model without any self-fulfilling prophecy indicator nor with controls for sample selection, all SEU parameters except the perceived costs of education C have a significant impact on students' educational success in terms of passing Abitur. Possibly, the fact that costs do not come into play in this model can be explained by the life course or selection hypothesis (Blossfeld and Shavit, 1993; Mare, 1980, 1993) which postulates that the effects of social inequality decrease during students' educational career. However, high school graduation still varies by the expected benefit of graduation, the expected amount of status decline and its expected impact, and the subjective probability of educational success.

Interestingly, when we introduce the SFP residuals from the performance model (*model 2a*), the latter are highly significant while the effects of B, $-SD$ and p_{sd} are canceled out, and the significance level of the estimate of p_{ep} drops down from the 99.9% level to

[19]The scale reliability of the IMR sum score is about Cronbach's $\alpha = .84$. Because the inter-correlations between the single IMR scores is in all cases near to one, the assumption of equal weights as it is always implied in simple sum scores is appropriate. For multicollinearity problems with lower inter-correlations a latent variable approach with free factor loadings for the IMR scores would be more adequate (Cohen et al., 2003). However, in our case, a confirmatory factor analysis with factor loadings around .99 (not shown) also strengthens the assumption of equal weights.

Table 2: Logistic Regression of Educational Success on SEU Predictors and Self-Fulfilling Prophecy Residuals

	Model 1a	*Model 2a*	*Model 3a*	*Model 4a*	*Model 1b*	*Model 2b*	*Model 3b*	*Model 4b*
	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value
B	1.21***	1.22	1.25*	1.29**	1.32**	1.21	1.24*	1.28**
	(3.54)	(1.94)	(2.21)	(2.80)	(3.17)	(1.86)	(2.09)	(2.64)
-SD	1.12*	1.21	1.23*	1.35**	1.31**	1.21	1.22	1.33**
	(2.05)	(1.82)	(1.96)	(3.05)	(2.76)	(1.71)	(1.87)	(2.86)
p_{sd}	0.88*	0.96	0.96	0.92	0.93	0.98	0.98	0.95
	(-2.18)	(-0.40)	(-0.40)	(-0.82)	(-0.79)	(-0.20)	(-0.21)	(-0.53)
p_{ep}	1.39***	1.30*	1.28*	1.46***	1.68***	1.30	1.29	1.48**
	(5.43)	(2.24)	(2.18)	(3.48)	(4.26)	(1.78)	(1.78)	(3.02)
C	0.96	1.00	1.00	0.98	1.02	1.02	1.03	1.02
	(-0.75)	(-0.02)	(0.01)	(-0.19)	(0.16)	(0.21)	(0.24)	(0.20)
r_{perf}		2.87***				2.85***		
		(10.47)				(9.97)		
r_{mot}			2.80***				2.78***	
			(10.26)				(9.99)	
r_{full}				1.84***				1.82***
				(6.37)				(6.07)
IMR_{sum}					1.16	0.67	0.65	0.83
					(1.33)	(-0.60)	(-0.68)	(-0.25)
IMR_{perf}						0.62		
						(-0.72)		
IMR_{mot}							0.60	
							(-0.79)	
IMR_{full}								0.74
								(-0.39)
Nagelkerke's R^2	0.06	0.36	0.35	0.21	0.12	0.35	0.35	0.21
N	1419.00	585.00	582.00	580.00	586.00	584.00	581.00	579.00

All coefficients are standardized odds ratios. Z-values in parentheses. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).
Variables: B: expected benefit; $-SD$: expected status decline; p_{sd}: expected impact of status decline; p_{ep}: expected success probability; r_{perf}: residuals (performance model); r_{mot}: residuals (motivation model); r_{full}: residuals (full model); $IMR_{sum}, IMR_{perf}, IMR_{mot}, IMR_{full}$: Inverse Mill's Ratios (control terms for sample selection bias).

the 95% level. In our theoretical section we have argued that $p_{ep} = f(SFP)$ (8), and although, admittedly, we are not able to model this impact over time, the drop-down in both effect size and significance of p_{ep} may strengthen this proposition. Moreover, for the case of the operationalization of self-fulfilling prophecies according to the performance model, we can conclude that they have a significant impact on students' educational success of passing Abitur: With regard to content, the probability to graduate immediately on the first try and with no class repetition is almost 2.9 times as high for students who have been over-estimated by their teachers with regard to their 10^{th} class academic performance compared to students who have been under-estimated.

In *model 3a* we can see that the residuals from the motivation model of table 1 also significantly predict students' educational success. However, and in line with the low predictive power of the motivation model, both effect size and t-value are lower compared to model 2a. Therefore, a teacher's evaluation nearly as a *thing in itself*, i.e. with no significant reduction in variation caused by its predictors, also significantly affects students' educational success. Yet this effect is enlarged if we are able to control for substantial over- and underestimations. In opposition to model 2a, the subjective expected benefit and the expected amount of status decline remain significant in *model 3a*.

Model *model 4a* shows that the residuals of the full model containing both performance and motivation predictors not only have a lower estimate and t-value compared to models 2a and 3a, but also lead to a decrease in model fit. Thus, if a teachers' evaluation is controlled for both students' performance and their motivation, over- and underestimations explain less of the variance of students' academic success. Moreover, if students' motivation is considered, differences in 'conventional' SEU parameters remain important.

When controls for sample selection are introduced in *Models 1b-4d*, the main difference to the a-models is that in two models, p_{ep} looses its significant impact on students' educational success. However, it is important to note that none of the self-fulfilling prophecy residuals is affected by sample selection correction. Since we indexed the Inverse Mills Ratios for the SEU predictors, we assume that this particular robustness is not an artefact of multicollinearity. Hence, while at least the distribution of success expectations may be explained by issues of social selectivity, it appears that for the case of educational success, the impact of over- and underestimations remains stable against social selectivity.

starting academic studies In table 3, the impact of both SEU predictors and self-fulfilling prophecy residuals on students' transitions to university are presented. Additionally to table 2, and according to the model by Becker (2003), the analyses also control for parental social class and education.

In contrast to table 2, in *model 5a*, only the expected benefit, B, and the subjective expected utility of successfully completing the chosen school track, p_{ep}, have a positive impact on student' propensity to start academic studies. Both indicators for the expected status decline as well as expected costs of education remain insignificant.

Table 3: Logistic Regression of Transition to University on SEU Predictors and Self-Fulfilling Prophecy Residuals

	Model 5a e^{b*sd}/z-value	*Model 6a* e^{b*sd}/z-value	*Model 7a* e^{b*sd}/z-value	*Model 8a* e^{b*sd}/z-value	*Model 5b* e^{b*sd}/z-value	*Model 6b* e^{b*sd}/z-value	*Model 7b* e^{b*sd}/z-value	*Model 8b* e^{b*sd}/z-value
B	1.17**	1.11	1.14	1.19	1.16	1.02	1.02	0.76
	(2.88)	(1.11)	(1.38)	(1.91)	(0.40)	(0.10)	(0.10)	(-0.95)
-SD	1.04	1.07	1.08	1.18	1.10	0.99	0.99	0.75
	(0.74)	(0.66)	(0.81)	(1.78)	(0.31)	(-0.08)	(-0.08)	(-0.98)
p_{sd}	0.92	1.06	1.06	1.03	1.03	1.07	1.07	1.11
	(-1.45)	(0.60)	(0.58)	(0.32)	(0.24)	(0.65)	(0.64)	(0.92)
p_{ep}	1.37***	1.40*	1.39*	1.58***	1.53	1.20	1.20	0.69
	(5.02)	(2.52)	(2.48)	(3.57)	(0.59)	(0.65)	(0.63)	(-0.76)
C	1.01	1.00	1.00	1.01	1.01	0.99	0.99	0.96
	(0.18)	(0.02)	(0.04)	(0.09)	(0.09)	(-0.11)	(-0.11)	(-0.39)
r_{perf}		2.53***				1.42		
		(9.42)				(0.38)		
r_{mot}			2.46***				1.42	
			(9.19)				(0.37)	
r_{full}				1.65***				0.59
				(5.50)				(-0.82)
social class	0.99	1.01	1.01	1.04	1.03	0.97	0.97	0.84
	(-0.21)	(0.11)	(0.09)	(0.30)	(0.12)	(-0.21)	(-0.20)	(-1.04)
education	1.12	1.07	1.07	1.11	1.14	1.07	1.07	1.11
	(1.63)	(0.53)	(0.55)	(0.87)	(1.17)	(0.51)	(0.49)	(0.83)
IMR_{1b}					0.85			
					(-0.17)			
IMR_{2b}						0.51		
						(-0.60)		
IMR_{3b}							0.51	
							(-0.59)	
IMR_{4b}								0.18
								(-1.59)
Nagelkerke's R^2	0.04	0.29	0.28	0.16	0.10	0.29	0.29	0.17
N	1414.00	584.00	581.00	579.00	579.00	579.00	579.00	579.00

All coefficients are standardized odds ratios. Z-values in parentheses. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$). Variables: B: expected benefit; $-SD$: expected status decline; p_{sd}: expected impact of status decline; p_{ep}: expected success probability; r_{perf}: residuals (performance model); r_{mot}: residuals (motivation model); r_{full}: residuals (full model); IMR_{1a} – $-IMR_{4a}$: Inverse Mill's Ratios (control terms for sample selection bias).

As in table 2, the effects for B are canceled out when self-fulfilling prophecies are introduced in *models 6a - 8a* – while the coefficient of p_{ep} remains stable. Again, the estimate for the performance-model residuals (*model 6a*) has the highest impact on the dependent variable, and the estimate for the full-model residuals the lowest (*model 8a*). However, compared to the estimates in models 1a-4a, the effect sizes diminished between 11.5 (full model) and 13.8 (motivation model) percent, and also the R^2 statistics are notably lower now. Thus, the effects of self-fulfilling prophecies seem to decrease in the educational life course.[20] Both parental social class and education do not have a significant effect, and the results did not differ when either one were removed from the models (not shown, available upon request).

Finally, in *models 5b - 8b*, we replicated models 5a to 8a with controls for sample selection. Therefore, we re-estimated the models from tables 2 in a probit equation (not shown, available upon request), stored the estimates as Inverse Mill's Ratios and included them in our models from tables 5. Although none of the Inverse Mill's Ratio coefficients in table 2 was significant, controlling for them affected the z-statistics of p_{ep}. Thus, to achieve more conservative estimates in the second-stage selection equation, we also controlled for the Inverse Mill's ratios from the first-stage selection equation.

Note that in table 3, each *model 5b-8b* is associated its own IMR_{1a} - IMR_{4a}. The results show that although the IMR scores themselves are not significant, they are able to cancel out the significant effects of both B and p_{sd} as well as those of the three residuals from models 5a – 8a. Hence, while in the case of the prediction of students' probability of educational success only one 'conventional' SEU predictors suffered from sample selection bias, if students' propensity of university transitions is controlled for the selectivity of the subsample, also the estimates of teachers' over- and under-estimations lack statistical significance.[21]

4.3 Sensitivity Analyses

One objection that could be made against the antecedent Heckit models (and, likewise, also against the models of Becker 2003) addresses the predictors in the selection equations. Particularly in the second-step selection equation (passing Abitur), the Inverse Mill's Ratios that had been stored from the first step might not suffice the exclusion restriction (for a similar line of arguing cf. Jürges and Schneider, 2006) for the third-step equation of interest (transition to university): Recall that in the third step, parental social class is again introduced as a covariate in the logit equation - while it had already been used as an instrument to identify the selection equation in the first step. Hence,

[20]Lucas (2001) suggests to rely on *predicted probabilities* rather than on regression coefficients when comparing changes of social background effects in the educational life course. Table D (Appendix) indicates that the above trend also holds for the predicted probabilities of high school graduations and university transitions, respectively (see corrected model).

[21]Since the number of observations for the two model sets with and without sample selection equation are not equal, we repeated the analyses for models 5a-8a without the observations who did not have a valid value for the Inverse Mill's Ratios. The results of the self-fulfilling prophecy residuals are robust against these modifications (not shown, available upon request).

the problem might arise that the third-step equation of interest might not be identified because it includes a variable that also affects our instrument, i.e. the IMR control terms in the second-step selection equation.

A strategy for situations where issues like selection and unobserved heterogeneity might arise but good instruments are not available[22] has been proposed by Buis (2010). The basic idea of his proposition is that unobserved variables that might affect both the main independent variable and the dependent variables over several transition points are captured by a weighted sum of random variables $\epsilon = \gamma z$ which are approximated by a normal distribution. To reflect a variety of scenarios as regards the distribution of this random variable, different values for the standard deviation of ϵ are assumed. If $sd(\epsilon) = 0$, the assumption of unobserved heterogeneity is completely discarded – which is the standard case in conventional OLS (or logit/probit) regressions. The higher the standard deviation of ϵ, the stronger the effect that is allowed for unobserved heterogeneity. In our case, we will examine how the effects of all of our self-fulfilling prophecy residuals change with $sd(\epsilon) = 0, sd(\epsilon) = 0.5, sd(\epsilon) = 1$, and $sd(\epsilon) = 2$, respectively.

In more formal terms, the two-step Heckit estimations of section 4.2 are amended by a sequential logit model where the probability of university transitions is conditional on the subsample of those who have passed Abitur (for all the following see Buis 2010, chapter 7):

$$p_{1i} = \frac{exp(\alpha_1 + \beta_1 SEU + \gamma_1 SFP)}{1 + exp(\alpha_1 + \beta_1 SEU + \gamma_1 SFP)} \tag{14}$$

and

$$p_{2i} = \frac{exp(\alpha_2 + \beta_2 SEU + \gamma_2 SFP)}{1 + exp(\alpha_2 + \beta_2 SEU + \gamma_2 SFP)} \text{ if } pass_{1i} = 1 \tag{15}$$

Let $\Lambda(u) = \frac{exp(u)}{1+exp(u)}$ capture the general functional form of the sequential logit model. If now the weighted sum of unobserved covariates, ϵ, is introduced, the expected probabilities of passing the two transitions averaged over ϵ read:

$$E_\epsilon(Pr[y \in \{B, C\}|x, \epsilon]) = E_\epsilon(\Lambda(\beta_{01} + \beta_{11} + \gamma z)) \tag{16}$$

$$E_\epsilon(Pr[y \in \{B, C\}|x, \epsilon]) = E_\epsilon(\Lambda(\beta_{02} + \beta_{12} + \underbrace{\gamma z}_{\epsilon})) \tag{17}$$

In a second step, we can also relax the restriction that this unobserved covariate is not a confounding variable – meaning that it is not correlated with our main predictor of interest, i.e. the self-fulfilling prophecy residuals. Just as we can approximate the potential impact of the unobserved covariate on the outcome by assuming different values for the standard deviation of the random variable, we can also approximate the potential

[22] While educational economists have proposed instruments like students' birth quarter (Angrist and Krueger, 1991) or distance to university (Card, 2001) to control for unobserved heterogeneity when measuring the returns of education, we believe that for our hypothesized effect of teachers' expectations on students' educational opportunities, it is uneven more challenging to find a good instrument that does affect the former but *not* the latter due to the logic of a self-fulfilling prophecy.

impact of the unobserved covariate on the self-fulfilling prophecy residuals by assuming different values for the correlation ρ between the two variables.

We think that by this technique we not only control for selectivity issues that may arise in a sequence of educational transitions, but also tackle the objection of 'private information' that may be part of the teachers' evaluation heuristic without being reflected in our self-fulfilling prophecy residuals. Since the scenarios that are simulated for the weighted sum of unobserved covariates "control" for a possible correlation with a specified independent variable, too, it is also possible to get an intuition about the direction that this private information may take.

To estimate these models, we would have to integrate over the distributions of $\gamma z (= \epsilon)$ – which requires numerical approximation using maximum simulated likelihood (Train, 2003) since there are no closed form solutions for the respective integrals (Buis, 2010). Luckily, this procedure has already been implemented in the `seqlogit` package (Buis, 2007) in Stata (StataCorp, 2009).

Figure 7 presents the sensitivity analyses for all three self-fulfilling prophecy residuals at both transition points. Each point on each single line represents a separate equation. For instance, the vertical axis of the line plot on the upper left show the parameter estimates and their 95% confidence intervals of the three residuals for students' probability on passing Abitur for the case that the correlation between the unobserved covariate and each residual would be zero. On the horizontal axis, however, these estimates are plotted against different assumed values for the standard deviation of the random variable that should approximate the unobservables, namely $sd = 0$ (the case of no unobserved covariate), $sd = 0.5$, $sd = 1$, and $sd = 2$. In the other three plots in the first line of the graph, the restriction of no correlation between the unobservable and the residuals is subsequently relaxed unto a correlation of $\rho = 0.5$. In the plots in the second line of the graph, this procedure is repeated for the estimates (and their 95% confidence intervals) of students' propensity of university transitions – conditional on having achieved Abitur before. Hence, the plots in figure 7 are based on $3 * 4 * 4 * 2 = 96$ equations in total: three for each residual, four for each standard deviation, four for each value of ρ, and two for each transition point.[23] If we start with the first set of equations predicting students' probability of passing Abitur, we see that in case of a zero or low (0.1) correlation between the unobservables and the residuals, an unobservable that affects the outcome might lead to an increase in the coefficients of the self-fulfilling prophecy residuals. Only if both the impact of the unobservable on the outcome and its correlation with the residuals are relatively strong, it might lower the estimates of the latter and likewise decrease their significance.

The same tendency holds for the parameter estimates for students propensity of university transitions. Just like in the selection model of table 8, the results lack statistical significance for the case of $\rho = 0$ and $= 0$. If the correlation between ϵ and the residuals is not too large, an increase in its standard deviation could be associated with an increase in the parameter estimates which may shift their confidence intervals above or

[23]The coefficients and z-statistics of the variables of interest for these equations are listed in tables E (Abitur) and F (university transitions) in the appendix.

next to the 95% significance level. However, if both the impact of the unobservable on the outcome and its correlation with the residuals are relatively strong again, it might lead to a decrease of both the estimates and their significance levels again. However, in that case the model would surely suffer from multicollinearity which would hinder an unambigous interpretation (Farrar and Glauber, 1967).

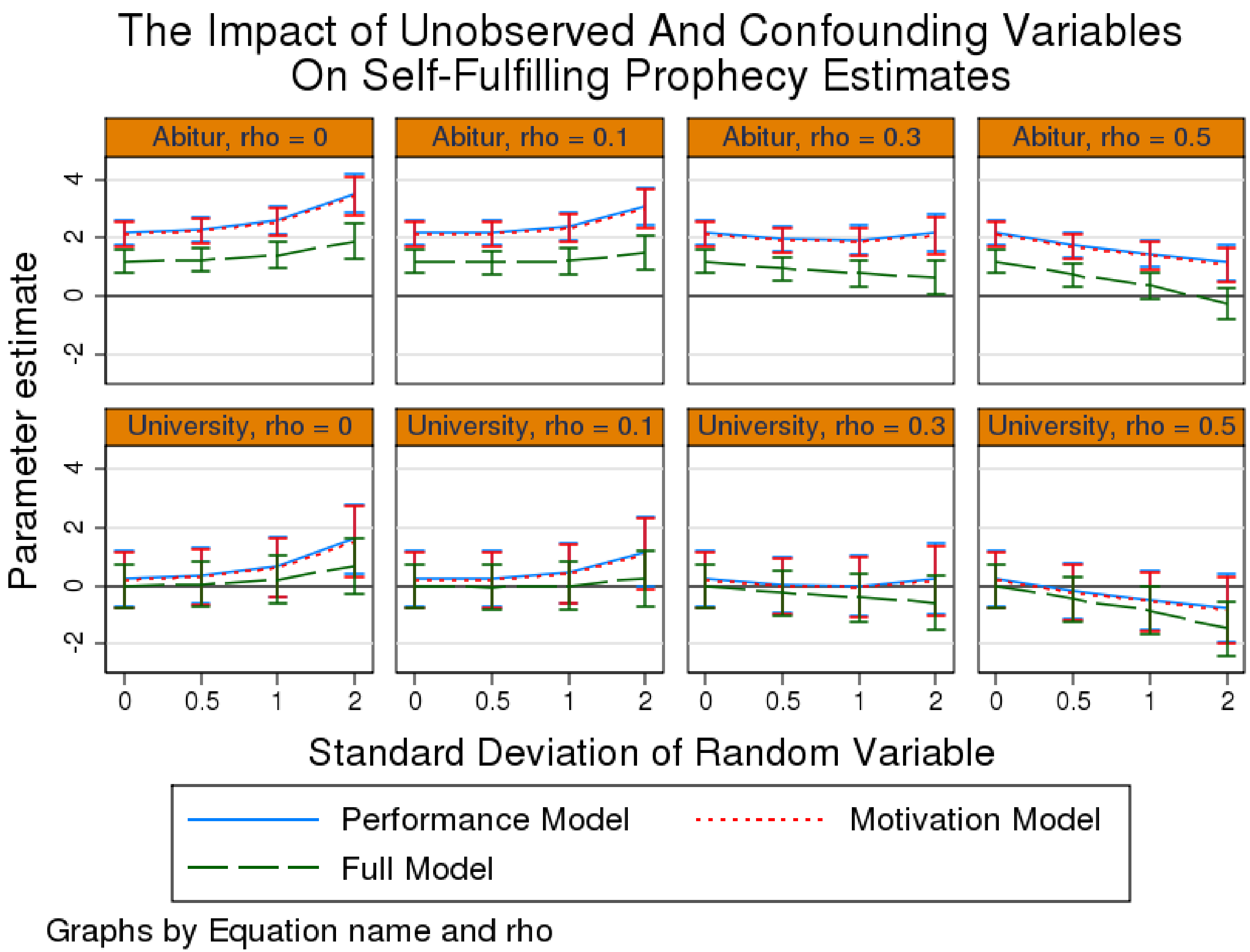


Figure 7: Sensitivity Analyses of Self-Fulfilling Prophecy Residuals

In sum, we can conclude that the prerequisites for an unobserved variable to weaken the estimates and/or significance levels of the self-fulfilling prophecy residuals have to be relatively strong. Neither the strength of its impact on the outcome nor its correlation with the variables of interest is a sufficient condition for decreasing the predictive power of the latter. Only if both conditions hold unto a relatively large extend, the results would not be robust. Since we would expect this to be an issue of multicollinearity, we do not consider the latter phenomenon to thwart our main findings.

5 Conclusion

The objective of this paper was to provide both theoretical and empirical evidence for the distinct effect of self-fulfilling prophecies apart from the conventional subjective-expected-utility (SEU) model of inequality in educational opportunities (IEO). Our aim was first to develop a formal model, and second to test this model to predict students' probability to graduate from high school (*Abitur*) as well as their subsequent university transitions.

In the theoretical section, we first summarized the basic assumptions of the SEU-IEO models by Breen and Goldthorpe (1997) and Esser (1999). After a literature review of *Pygmalion* and self-fulfilling prophecy research (Rosenthal and Jacobson, 1968; Madon et al., 1997; Jussim and Harber, 2005), we brought the argument that its main finding, i.e. that teachers' expectations may influence students' academic performance, requires an extension of the present SEU-IEO model. Consequently, an integration of self-fulfilling prophecies in the formal SEU-IEO model by Esser (1999) was developed.

Methodologically, self-fulfilling prophecies were operationalized as the residuals of a regression of a specific form of teachers' evaluations on a performative and a motivational set of variables (also see Madon et al., 1997). However, in the empirical section it turned out that the performance model was able to predict teachers' evaluations more satisfactory than the motivation model.

In our multivariate analyses that were based on the "Kölner Gymnasiasten-Panel", we found that the predictive power of the conventional SEU-IEO model is by average weaker than in previous studies (e.g. Becker, 2003; Becker and Hecken, 2009) – which could be a corroboration of the life-course hypothesis (Mare, 1980, 1993) that indicates that the effects of social inequality decrease during students' educational career.

In contrast, at least in the baseline model, the self-fulfilling prophecy residuals were able to predict both students' educational success in terms of passing Abitur and their university transitions significantly. Thus, the tentative conclusion from these models would be that self-fulfilling prophecies have indeed distinct effects besides the conventional SEU predictors. Moreover, since the effect sizes of the residuals are lower for students' university transitions than for their educational success, this could be another demonstration of life course effects.

Because of the selectivity problem concerning the dependency of students' resources and preferences by social class and their conditional transition decisions, we replicated all models with corrections for sample selection bias (Heckman, 1979). It turned out that in case of the prediction of students' educational success the results remain stable, while in case of the prediction of their university transitions all self-fulfilling prophecy residuals lost their significance. This indicates that there is little evidence that the efficacy of self-fulfilling prophecies could mainly be explained by students' social class. Notwithstanding this particular stability, there is no reason to assume that self-fulfilling prophecies might affect students' propensity of university transitions *conditional on having passed Abitur.* This suggests that the effect of teachers' expectations is limited on students' school success and does not influence their decision for or against starting academic studies.

Because of some methodological objections that could be made against the quality of the instruments in the selection models, and to tackle the argument that teachers might have private information at their disposal which is not captured by the variables in the three residual models, a sensitivity analysis was performed. In particular, we additionally allowed for unobserved heterogeneity which was approximated by a random variable that could take different values on both its standard deviations and its correlation with the self-fulfilling prophecy residuals. It turned out that only if relatively high values on both parameters are allowed simultaneously, the residual estimates might not be robust. However, since this would go in line with the problem of multicollinearity, we do not expect our main findings to be challenged by this issue.

Nonetheless, further analyses should consider additional variables. Remember that one major theoretical shortcoming of *Pygmalion* concerns an insufficient consideration of moderators such as students' grade level or teachers' duration in class. Thus, future studies should also include potential covariates *beyond* the standard SEU predictors to ensure a better understanding of the social mechanism behind the efficacy of self-fulfilling prophecies – particularly, if the empirical model *par se* is, as in the case of our data at hand, only an approximation of the theoretical or *substantive* model. Considering both teacher- and student-level variables would require to estimate a cross-classified hierarchical model (Snijders and Bosker, 1999; Hox, 2002) wherein teachers' evaluations as the lowest unit are nested in both teacher and student contexts. To be sure, this might also necessitate a refined operationalization of self-fulfilling prophecies.

References

Allison PD (2002) *Missing Data.* Thousand Oaks, CA: Sage.

Amthauer R (1957) *Intelligenz-Struktur-Test.* Göttingen: Hogrefe.

Angrist JD and Krueger AB (1991) Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics* 106(4): 979–1014.

Ashenfelter O; Harmon C and Oosterbeek H (1999) A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour economics* 6(4): 453–470.

Becker R (2000) Klassenlage und Bildungsentscheidungen - Eine empirische Anwendung der Wert-Erwartungstheorie. *Zeitschrift für Soziologie* 52(3): 450–474.

Becker R (2003) Educational Expansion and Persistent Inequalities of Education. *European Sociological Review* 19(1): 1–24.

Becker R and Hecken AE (2008) Why are Working-class Children Diverted from Universities?–An Empirical Assessment of the Diversion Thesis. *European Sociological Review* 25(2): 233–250.

Becker R and Hecken AE (2009) Higher Education or Vocational Training?: An Empirical Test of the Rational Action Model of Educational Choices Suggested by Breen and Goldthorpe and Esser. *Acta Sociologica* 52(1): 25–45.

Bishop JH (1989) Is the test score decline responsible for the productivity growth decline? *The American Economic Review* 79(1): 178–197.

Blossfeld HP and Shavit Y (1993) *Persisting Barriers*, 1–23. Boulder, CO: Westview Press.

Boissiere M; Knight JB and Sabot RH (1985) Earnings, schooling, ability, and cognitive skills. *The American Economic Review* 75(5): 1016–1030.

Boudon R (1974) *Education, Opportunity and Social Inequality: Changing Prospects in Western Society.* New York: John Wiley & Sons.

Brattesani KA; Weinstein RS and Marshall HH (1984) Student Perceptions of Differential Teacher Treatment as Moderators of Teacher Expectation Effects. *Journal of Educational Psychology* 76(2): 236–247.

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

Breen R and Jonsson JO (2000) Analyzing educational careers: A multinomial transition model. *American Sociological Review* 754–772.

Buis ML (2007) SEQLOGIT: Stata module to fit a sequential logit model.
URL: *http://ideas.repec.org/c/boc/bocode/s456843.html*

Buis ML (2010) Inequality of Educational Outcome and Inequality of Educational Opportunity in the Netherlands during the 20th Century. Disseration thesis, Vrije Universiteit Amsterdam.

Card D (2001) Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69(5): 1127–1160.

Cohen J; Cohen P; West SG et al. (2003) *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum.

Cox DR (1990) Role of models in statistical analysis. *Statistical Science* 5(3): 169–174.

Cunha F and Heckman JJ (2007) Identifying and estimating the distributions of ex post and ex ante returns to schooling. *Labour Economics* 14(6): 870–893.

Cunha F; Heckman JJ and Navarro S (2005) Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers* 57(2): 191–261.

Elashoff JD and Snow RE (1971) *Pygmalion reconsidered: : A case study in statistical inference: Reconsideration of the Rosenthal-Jacobson data on teacher expectancy.* Worthington, Ohio: CA Jones Pub. Co.

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

Farrar DE and Glauber RR (1967) Multicollinearity in regression analysis: The problem revisited. *The Review of Economic and Statistics* 49(1): 92–107.

Gelman A and Hill J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Goldthorpe JH (1996) Class analysis and the reorientation of class theory: the case of persisting differentials in educational attainment. *British Journal of Sociology* 47(3): 481–505.

Goldthorpe JH (2001) Causation, statistics, and sociology. *European Sociological Review* 17(1): 1–20.

Greene WH (2003) *Econometric analysis.* Upper Saddle River, NJ: Prentice Hall.

Hanushek EA and Kimko DD (2000) Schooling, labor-force quality, and the growth of nations. *The American Economic Review* 90(5): 1184–1208.

Heckman JJ (1979) Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* 47(1): 153–161.

Hillmert S and Jacob M (2010) Selections and social selectivity on the academic track: A life-course analysis of educational attainment in Germany. *Research in Social Stratification and Mobility* 28(1): 59–76.

Hox JJ (2002) *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Jensen AR (1969) How much can we boost IQ and scholastic achievement. *Harvard Educational Review* 39(1): 1–123.

Jonsson JO (1999) Explaining Sex Differences in Educational Choice An Empirical Assessment of a Rational Choice Model. *European Sociological Review* 15(4): 391–404.

Jürges H and Schneider K (2006) Age at school entry and teacher's recommendations for secondary school track choice in Germany. In: *Paper presented at the 20th annual ESPE conference, Verona*.

Jussim L; Eccles J and Madon SJ (1996) *Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy*, vol. 28, 281–388. San Diego: Academic Press.

Jussim L and Harber KD (2005) Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies. *Personality and Social Psychology Review* 9(2): 131–155.

Lauen DL (2007) Contextual Explanations of School Choice. *Sociology of Education* 80(3): 179–209.

Lucas SR (2001) Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects. *American Journal of Sociology* 106(6): 1642–1690.

Madon S; Jussim L and Eccles J (1997) In Search of the Powerful Self-Fulfilling Prophecy. *Journal of Personality and Social Psychology* 72(4): 791–809.

Madon S; Jussim L; Keiper S et al. (1998) The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin* 24(12): 1304–1318.

Mare RD (1980) Social Background and School Continuation Decisions. *Journal of the American Statistical Association* 75(370): 295– 305.

Mare RD (1993) *Educational stratification on observed and unobserved components of family background*, 351–376. Boulder, CO: Westview Press.

Merton RK (1948) The self-fulfilling prophecy. *Antioch Review* 8(2): 193–210.

Müller-Benedict V (2007) Wodurch kann die soziale Ungleichheit des Schulerfolgs am stärksten verringert werden? *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59(4): 615–639.

Need A and De Jong U (2001) Educational differentials in the Netherlands. *Rationality and Society* 13(1): 71–98.

Pietsch M and Stubbe TC (2007) Inequality in the transition from primary to secondary school: School choices and educational disparities in Germany. *European Educational Research Journal* 6(4): 424–445.

Raudenbush SW (1984) Magnitude of Teacher Expectancy Effects on Pupil IQ as a Function of the Credibility of Expectancy Induction: A Synthesis of Findings From 18 Experiments. *Journal of Educational Psychology* 76(1): 85–97.

Raudenbush SW (1994) *Random effects models*, 301–321. New York: Sage.

Rogosa D (1987) Casual models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics* 12(2): 185–195.

Rosenthal R and Jacobson L (1968) *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart & Winston.

Rosenthal R and Rubin DB (1978) Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences* 1(3): 377–415.

Roth BM (1995) We can throw teacher expactations on the IQ scrap heap. New York Times.

Schelsky H (1957) *Schule und Erziehung in der industriellen Gesellschaft*. Würzburg: Werkbund Verlag.

Schneider T (2008) Social Inequality in Educational Participation in the German School System in a Longitudinal Perspective: Pathways into and out of the most Prestigious School Track. *European Sociological Review* 24(4): 511–526.

Smith ML (1980) Meta-analysis of research in teacher expectation. *Evaluation in Education* 4: 53–56.

Snijders TAB and Bosker RJ (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications Ltd.

Snow RE (1969) Unfinished pygmalion. *Contemporary Psychology* 14(4): 197–199.

Sø rensen AB (1998) *Theoretical mechanisms and the empirical study of social processes*, 238–266. Cambridge: Cambridge University Press.

StataCorp (2009) *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Stocké V (2007) Explaining Educational Decision and Effects of Families' Social Class Position: An Empirical Test of the Breen Goldthorpe Model of Educational Attainment. *European Sociological Review* 23(4): 505–519.

Thorndike RL (1968) Review of Pygmalion in the classroom. *American Educational Research Journal* 5(4): 708–711.

Train KA (2003) *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Wineburg SS (1987) The Self-Fulfillment of the Self-Fulfilling Prophecy. *Educational Researcher* 16(9): 28–37.

Wooldridge J (2006) *Introductory Econometrics*. Cincinnati, Ohio: South-Western College Publ.

6 Appendix

Table A: Descriptive Results

	(1)				
	count	mean	sd	min	max
time of high school graduation	1415	81.90	14.66	46.00	228.00
graduation on first try	1987	0.53	0.50	0.00	1.00
time of university transition	1987	154.57	108.17	44.00	367.00
university transition within 3 years after graduation	1987	0.48	0.50	0.00	1.00
teacher evaluation (dichotomized)	1367	0.55	0.50	0.00	1.00
intelligence	3230	110.45	11.35	76.00	151.00
average grade	3227	499.98	69.22	221.00	703.00
motivation	3224	5.71	2.08	1.00	11.00
self-concept	3208	6.65	1.91	1.00	11.00
self-confidence	3213	8.13	1.51	1.00	11.00
residuals (performance model)	1309	0.00	0.38	-0.99	0.99
residuals (motivarion model)	1294	0.01	0.44	-0.91	0.95
residuals (full model)	1287	0.00	0.38	-1.20	1.05
residuals (performance model, dichotomized)	1309	0.55	0.50	0.00	1.00
residuals (motivarion model, dichotomized)	1294	0.56	0.50	0.00	1.00
residuals (full model, dichotomized)	1287	0.48	0.50	0.00	1.00
B	3225	0.69	0.46	0.00	1.00
-SD	2355	0.41	0.49	0.00	1.00
Psd	2674	0.37	0.48	0.00	1.00
Pep	2695	0.95	0.23	0.00	1.00
C	2695	0.46	0.50	0.00	1.00
parental social class	2687	49.37	12.63	18.00	78.00
parental educational attainment	3374	2.14	1.23	1.00	4.00
educational motivation	2290	0.80	0.40	0.00	1.00
investment risk	2691	0.44	0.50	0.00	1.00
IMR_sum	926	4.22	0.45	3.44	5.33
IMR_perf_di	1070	0.69	0.10	0.47	0.94
IMR_mot_di	1058	0.68	0.10	0.47	0.92
IMR_full_di	1054	0.82	0.09	0.63	1.04
IMR_1a	1419	0.62	0.16	0.46	1.53
IMR_2a	585	0.62	0.44	0.24	1.79
IMR_3a	582	0.62	0.43	0.24	1.80
IMR_4a	580	0.60	0.32	0.24	1.82
IMR_1b	579	0.60	0.26	0.32	1.91
IMR_2b	579	0.62	0.43	0.21	1.81
IMR_3b	579	0.62	0.43	0.21	1.81
IMR_4b	579	0.60	0.32	0.22	1.89
IMR_1c	1419	0.61	0.06	0.58	0.75
IMR_2c	585	0.61	0.41	0.29	1.30
IMR_3c	582	0.61	0.40	0.29	1.30
IMR_4c	580	0.59	0.25	0.33	1.05
IMR_1d	579	0.59	0.12	0.40	0.94
IMR_2d	579	0.61	0.41	0.24	1.39
IMR_3d	579	0.61	0.41	0.24	1.39
IMR_4d	579	0.59	0.25	0.26	1.15

Table B: Logistic Regression of Educational Success on SEU Interaction Terms and Self-Fulfilling Prophecy Residuals

	Model 1c	*Model 2c*	*Model 3c*	*Model 4c*	*Model 1d*	*Model 2d*	*Model 3d*	*Model 4d*
	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value
$B + p_{sd} * SD$	1.19**	1.16	1.19	1.29**	1.29**	1.15	1.18	1.27**
	(3.20)	(1.50)	(1.78)	(2.85)	(3.01)	(1.43)	(1.70)	(2.68)
C/p_{ep}	0.98	1.05	1.05	1.05	1.13	1.09	1.09	1.11
	(-0.28)	(0.48)	(0.45)	(0.50)	(1.34)	(0.81)	(0.78)	(1.06)
r_{perf}		3.11***				3.07***		
		(11.59)				(11.44)		
r_{mot}			3.02***				2.99***	
			(11.35)				(11.20)	
r_{full}				1.98***				1.95***
				(7.36)				(7.18)
IMR_{sum}					1.22*	0.72	0.72	0.95
					(2.05)	(-0.50)	(-0.51)	(-0.07)
IMR_{perf}						0.66		
						(-0.65)		
IMR_{mot}							0.65	
							(-0.66)	
IMR_{full}								0.82
								(-0.30)
Nagelkerke's R^2	0.01	0.34	0.33	0.15	0.03	0.34	0.33	0.15
N	1419.00	585.00	582.00	580.00	586.00	584.00	581.00	579.00

All coefficients are standardized odds ratios. Z-values in parentheses. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).
Variables: $B + P_{sd} * SD$: educational motivation; C/p_{ep}: investment risk; r_{perf}: residuals (performance model); r_{mot}: residuals (motivation model); r_{full}: residuals (full model); $IMR_{sum}, IMR_{perf}, IMR_{mot}, IMR_{full}$: Inverse Mill's Ratios (control terms for sample selection bias).

Table C: Logistic Regression of Transition to University on SEU Interaction Terms and Self-Fulfilling Prophecy Residuals With Controls for Sample Selection

	Model 5c	*Model 6c*	*Model 7c*	*Model 8c*	*Model 5d*	*Model 6d*	*Model 7d*	*Model 8d*
	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value	e^{b*sd}/z-value
$B + p_{sd} * SD$	1.12*	1.04	1.07	1.16	1.00	0.85	0.85	0.70
	(2.16)	(0.44)	(0.70)	(1.68)	(-0.01)	(-0.84)	(-0.85)	(-1.14)
C/p_{ep}	1.06	1.10	1.10	1.13	1.07	0.97	0.97	0.90
	(1.03)	(0.91)	(0.89)	(1.23)	(0.39)	(-0.19)	(-0.18)	(-0.61)
r_{perf}		2.73***				0.48		
		(10.46)				(-0.53)		
r_{mot}			2.65***				0.48	
			(10.22)				(-0.55)	
r_{full}				1.77***				0.47
				(6.47)				(-0.96)
social class	1.01	1.03	1.02	1.05	0.96	0.89	0.89	0.79
	(0.09)	(0.22)	(0.18)	(0.47)	(-0.16)	(-0.69)	(-0.70)	(-1.17)
education	1.14	1.05	1.06	1.10	1.14	1.07	1.07	1.10
	(1.91)	(0.42)	(0.45)	(0.86)	(1.18)	(0.48)	(0.48)	(0.79)
IMR_{1d}					0.80			
					(-0.53)			
IMR_{2d}						0.17		
						(-1.25)		
IMR_{3d}							0.17	
							(-1.30)	
IMR_{4d}								0.22
								(-1.67)
Nagelkerke's R^2	0.01	0.27	0.26	0.12	0.02	0.27	0.27	0.12
N	1414.00	584.00	581.00	579.00	579.00	579.00	579.00	579.00

All coefficients are standardized odds ratios. Z-values in parentheses. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$). Variables: B: expected benefit; $-SD$: expected status decline; p_{sd}: expected impact of status decline; p_{ep}: expected success probability; r_{perf}: residuals (performance model); r_{mot}: residuals (motivation model); r_{full}: residuals (full model); $IMR_{1a} - -IMR_{4a}$: Inverse Mill's Ratios (control terms for sample selection bias).

Table D: Predicted Probabilities of High-School Graduations and University Transitions

		High School Graduation		**University transition**		**Difference**	
	value	*uncorrected*	*corrected*	*uncorrected*	*corrected*	*uncorrected*	*corrected*
performance residuals	0	0,33	0,33	0,28	0,46	0,04	-0,13
	1	0,82	0,81	0,73	0,64	0,08	0,17
motivation residuals	0	0,33	0,34	0,29	0,46	0,04	-0,12
	1	0,81	0,81	0,73	0,64	0,08	0,17
full model residuals	0	0,50	0,51	0,45	0,70	0,05	-0,20
	1	0,77	0,77	0,69	0,45	0,09	0,32

Note: All values are predicted probabilities conditional on positive and negative values for each self-fulfilling prophecy residual, respectively. Corrected model: Includes Inverse Mill's Ratio controls for sample selection.

Table E: A Sensitivity Analysis for Self-Fulfilling Prophecy Residuals (Transition: Abitur)

	performance model	*motivation model*	*full model*
rho0sd0a			
b	2.184151	2.131192	1.195157
z	10.32542	10.10805	6.190385
rho0sd5a			
b	2.300063	2.243347	1.253663
z	10.3838	10.15775	6.189155
rho0sd10a			
b	2.604713	2.538721	1.409969
z	10.47301	10.23553	6.186296
rho0sd20a			
b	3.520772	3.429911	1.890575
z	10.5664	10.31637	6.179325
rho1sd0a			
b	2.184151	2.131192	1.195157
z	10.32542	10.10805	6.190385
rho1sd5a			
b	2.194397	2.137723	1.153021
z	9.911121	9.683768	5.694879
rho1sd10a			
b	2.391868	2.326	1.207879
z	9.63034	9.390714	5.30694
rho1sd20a			
b	3.092165	3.001605	1.484773
z	9.307114	9.054466	4.867169
rho3sd0a			
b	2.184151	2.131192	1.195157
z	10.32542	10.10805	6.190385
rho3sd5a			
b	1.976348	1.919971	.9483301
z	8.95787	8.728262	4.70098
rho3sd10a			
b	1.943463	1.878511	.7919539
z	7.912032	7.66861	3.518721
rho3sd20a			
b	2.171271	2.082975	.6395354
z	6.693621	6.435879	2.147459
rho5sd0a			
b	2.184151	2.131192	1.195157
z	10.32542	10.10805	6.190385
rho5sd5a			

	performance model	*motivation model*	*full model*
b	1.749196	1.693402	.7390304
z	7.985123	7.753716	3.690597
rho5sd10a			
b	1.463371	1.400289	.3597174
z	6.096067	5.84949	1.635921
rho5sd20a			
b	1.158377	1.074779	-.2541716
z	3.761082	3.497863	-.8991183

Table F: A Sensitivity Analysis for Self-Fulfilling Prophecy Residuals (Transition: University)

	performance model	*motivation model*	*full model*
rho0sd0b			
b	.231838	.1909521	-.0108028
z	.4816385	.3980592	-.0283692
rho0sd5b			
b	.3499028	.3060727	.0476244
z	.7094946	.6228028	.1221588
rho0sd10b			
b	.6600259	.6073403	.2016426
z	1.254707	1.15886	.4853785
rho0sd20b			
b	1.593731	1.511577	.6637602
z	2.577975	2.454781	1.351818
rho1sd0b			
b	.231838	.1909521	-.0108028
z	.4816385	.3980592	-.0283692
rho1sd5b			
b	.2442164	.2004237	-.053016
z	.4953126	.4079225	-.1360199
rho1sd10b			
b	.4471103	.3945533	-.0004007
z	.8506118	.7534233	-.0009652
rho1sd20b			
b	1.164963	1.083175	.2582188
z	1.887321	1.761783	.526771
rho3sd0b			
b	.231838	.1909521	-.0108028
z	.4816385	.3980592	-.0283692
rho3sd5b			
b	.0260053	-.0175311	-.2576976
z	.0528434	-.0357488	-.6623826

	performance model	*motivation model*	*full model*
rho3sd10b			
b	-.0018588	-.053469	-.4159646
z	-.0035585	-.1027417	-1.008306
rho3sd20b			
b	.24269	.1637122	-.5849299
z	.3982183	.2696962	-1.20981
rho5sd0b			
b	.231838	.1909521	-.0108028
z	.4816385	.3980592	-.0283692
rho5sd5b			
b	-.2014799	-.2445182	-.4669888
z	-.410981	-.5005141	-1.204813
rho5sd10b			
b	-.483083	-.5328065	-.847538
z	-.9369038	-1.037127	-2.081037
rho5sd20b			
b	-.7734081	-.8465689	-1.474437
z	-1.304974	-1.434111	-3.141754