Pygmalion's Long Shadow

Determinants and Outcomes of Teachers' Evaluations

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2012

vorgelegt
von

Dominik Becker, M.A.

aus

Köln

Referent: Prof. Dr. Heiner Meulemann
Korreferent: Prof. Dr. Rolf Becker
Tag der Promotion: 14.12.2012

Preface

The volume at hand comprises four papers – preceded by a comprehensive introduction – that are intended to obtain the doctoral degree to be awarded by the Faculty of Management, Economics and Social Sciences of the University of Cologne. This preface aims to clarify the publication status of the four papers.

The first paper, *Teachers' Evaluations and the Social Situation in the Classroom* (co-authored by Klaus Birkelbach) was submitted to *Sociology of Education*.

An earlier version of the second paper, *Intelligence and Academic Achievement as Contextual-Level Predictors of Teachers' Evaluations: Big-Fish-Little-Pond or Reflected-Glory Effect?* (co-authored by Klaus Birkelbach) appeared as a book chapter (*Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish- Little-Pond- oder Reflected-Glory-Effekt? Eine Mehrebenen- analyse von Lehrerurteilen*) in *Komparative Sozialforschung* (ed. by T. Beckers, K. Birkelbach, J. Hagenah and U. Rosar; Wiesbaden 2010: Springer). Although the contributions to that edited volume have already been peer-reviewed, the paper at hand extends its German predecessor both theoretically and empirically.

An earlier draft of the third paper was accepted for presentation as a full conference paper at both the conference "Higher education and beyond – Inequalities regarding entrance to higher education and educational credentials", July 5-9, 2010, Monte Verita, and at a RC04 (Research Committee of the International Sociological Association on Sociology of Education) poster session at the XVII ISA World Congress of Sociology, July 11-17, 2010, Gothenburg. A revised version was recently accepted at *Rationality and Society*. Comments by two anonymous referees of that journal are already considered in the version at hand.

The fourth paper, *Does the Effect of Teachers' Expectations on Students' Educational Opportunities Decrease over Educational transitions? A Statistical Matching Approach*, was accepted for presentation as a full conference paper both at the RC28 (Research Committee of the International Sociological Association on Social Stratification and Mobility) Spring Meeting, April 13-16, 2011, Essex, and at the fourth Conference of the European Survey Research Association (ESRA), July 18-22, 2011, Lausanne. Comments by participants of both conferences as well as of two anonymous ESRA reviewers have been considered.

In both co-authored papers, I am first and corresponding author.

Acknowledgements First and foremost, I am grateful to my co-author and mentor at the Cologne Graduate School in Management, Economics and Social Sciences, Klaus Birkelbach, to my two supervisors Heiner Meulemann and Rolf Becker, and to my colleagues Wiebke Breustedt, Pascal Siegers and Christina Zuber for their outstanding academic inspiration and support.

And finally, I would like to thank all members of the Cologne Graduate School in Management, Economics and Social Sciences, all members of the SOCLIFE graduate school, all participants in the colloquium of the Chair of Sociology of the University of Cologne, and all participants and reviewers of the above-mentioned conferences and journals for their valuable comments on earlier versions of the papers.

Cologne, August 2013

Table of Contents

I. Causes and Effects of Teachers' Evaluations: A Theoretical Primer

1 Introduction

> "This theory design pushes the presentation to unusually high levels of abstraction. Our flight must take place above the clouds, and we must reckon with a rather thick cloud cover. We must rely on our instruments. Occasionally, we may catch glimpses below of a land with roads, towns, rivers, and coastlines that remind us of something familiar, or glimpses of a larger stretch of landscape with the extinct volcanoes of Marxism. But no one should fall victim to the illusion that these few points of reference are sufficient to guide our flight" (Luhmann, 1995 [1984], foreword to the German edition, p. l).

Since more than half a century, and in both public debates and scientific discourses, the idea of meritocracy (Young, 1958) more or less serves as a benchmark according to which an educational system's justice and effectiveness were to be judged. However, since several decades, educational sociology unveils both theoretically and empirically that educational systems are neither perfectly fair (Coleman, 1966; Jencks, 1972; Blossfeld and Shavit, 1993), nor the idea of meritocracy is incontestable regarding its moral implications (Bell, 1972; Goldthorpe, 1996b; Solga, 2005; Becker and Hadjar, 2011).

While in the beginning of educational sociology, *i*nequalities in *e*ducational *o*pportunities (IEO) were typically explained by the postulate that the value of education *per se* varies by social strata (Hyman, 1953), more recent (and also more parsimonious) theoretical accounts discard this demanding hypothesis in favor of only assuming the underlying *cost-benefit* considerations of an educational transition decision to be class-variant – while parents' appraisement of education *per se* (i.e. in an absolute rather than in a relative sense) could remain constant (Keller and Zavalloni, 1964; Boudon, 1974; Meulemann, 1979; Goldthorpe, 1996a; Erikson and Jonsson, 1996a; Breen and Goldthorpe, 1997; Esser, 1999).

On the other hand, Wisconsin status attainment theorists (Sewell et al., 1969, 1970) have already brought up students' *significant others* as an important variable affecting their educational outcomes including their aspirations (also see Morgan, 2006). One prominent example of these significant others are teachers whose expectations were shown to affect both student academic self-concept and achievement in various *Pygmalion* and self-fulfilling prophecy studies (Rosenthal and Jacobson, 1968; Raudenbush, 1984; Jussim, 1986; Madon et al., 1997; Jussim and Harber, 2005).

Yet, Wisconsin status attainment theorists have been seriously criticized for confounding the correlations obtained in their path models with causality (Freedman, 1987; Hedström and Swedberg, 1996; Hedström, 2005; Hedström and Ylikoski, 2010). Furthermore, I see still need for a more thorough theoretical specification concerning the *generative processes* (Goldthorpe, 2001) of teacher expectancy effects. This is the gap I intend to close with the volume at hand in general, and by means of its introduction in particular.

This volume comprises two papers analyzing the predictors of teachers' evaluations, and another two with the latter's outcomes as the crucial objective. Concretely, in the data at hand (the *Cologne High School Panel*), teachers had been asked whom of their 10^{th} class students they consider to be able to start academic studies, and whom of them not. The first paper models these evaluations as an outcome of students' cognitive ability in terms of intelligence scores, their average grades, their parents' social class, and their aspirations. Using structural equation modeling as the method of analysis, the paper's proximity to the Wisconsin status attainment tradition and thus the need for a solid theoretical foundation is evident.

The second paper adds another level of analysis by investigating to what extent teachers' evaluations depend on reference-group effects in the classroom. While the techniques of multilevel analysis provide a sophisticated statistical framework for testing contextual-level hypotheses (Bryk and Raudenbush, 1992; Snijders and Bosker, 1999; Gelman and Hill, 2007; Hox, 2010), multilevel *theory* was quite prominent in the early 1970s, then slowly abated until entirely stagnating since the early 1990s (Hauser, 1970a; Barton, 1970; Hauser, 1970b; Farkas, 1974; Hauser, 1974; Blalock, 1984; Van den Eeden, 1992). Therefore, I aim to discuss how these theoretical consideration can be connected to contemporary reference-group effect research on teachers' evaluations.

The third paper asks how self-fulfilling prophecy effects of teachers' expectations – measured by their evaluations – relate to cost-benefit-based theories about social inequality in educational opportunities. As mentioned above, in my view, self-fulfilling prophecy research still stands in the duty of a more fine-grained specification of how the implied teacher treatment effect can be assumed to affect student achievement. Below, I will argue that this effect should be understood as operating via *students' subjective expected probability of educational success*, as it is referred to in IEO research (Breen and Goldthorpe, 1997; Esser, 1999).

And finally, the fourth paper analyzes to what extent the above self-fulfilling prophecy effect might vary over a sequence of educational transitions. More precisely, this question condenses to the problem of how students' beliefs in terms of their subjective expected probability of educational success change via a mechanism of *Bayesian updating* conditional on having successfully passed a preceding transition (Breen, 1999; Breen and García-Peñalosa, 2002).

While the title of this volume merely addresses the *Determinants and Outcomes of Teachers' Evaluations*, in this introduction, I intend to get closer to the underlying *causes and effects*. The theoretical framework whereby this should be accomplished is the one of *social mechanisms* (Elster, 1985, 1989; Hedström and Swedberg, 1996, 1998b; Gambetta, 1998; Hedström, 2005). The crucial objective of this theoretical account is to get close to an *understanding explanation* in the Weberian sense *by opening the* *black box and showing the cogs and wheels of the machinery* (Elster, 1985, 1989). It is argued that both the covering-law tradition (Hempel, 1942; Hempel and Oppenheim, 1948; Hempel, 1965; see Opp, 2005b for a still prominent textbook example) and what has been referred to as *variable sociology* (Esser, 1999) or as the *robust dependence tradition* (Goldthorpe, 2001) in sociology fall behind this demand: first, by overhasty postulating the existence of general laws in the social sciences without trying to understand actors' motives, beliefs and means; and second, by confounding correlations with causality without providing a sufficient theoretical justification for the implied causal structure. In contrast, mechanism-based explanations build on what Popper (1945a, 1994) called the analysis of actors' *social situation*, i.e., the reconstruction of both actors' external restrictions and their motives and beliefs by means of a suitable theory of action. As I will argue below, there is a surprisingly consistent line of reasoning from Weber's demand for an *understanding* explanation (and his concept of *Richtigkeitsrationalität*) over Popper's situational logic combined with its *rationality principle* on to contemporary mechanism-based explanations distinguishing between *desire-mediated*, *belief-mediated*, and *opportunity-mediated* social mechanisms (Hedström and Swedberg, 1996, 1998b; Hedström, 2005; Hedström and Ylikoski, 2010). These latter dimensions can also be used as a starting point for a comparatively weak rational action theory.

Hence, this introductory chapter provides a theoretical foundation of the underlying social mechanisms regarding both IEO research and the determinants and outcomes of teachers' evaluations analyzed in the four papers of the volume at hand. I will argue that following the Keller and Zavalloni (1964) and Boudon (1974) tradition of discarding the assumption of a class-dependent *absolute* value of education, differences in educational aspirations due to differences in cost-benefit considerations can mainly be accounted for by *belief-mediated* mechanisms. The same holds for both teachers' action scripts that shape their evaluations (cf. paper 1) and teacher expectancy effects (in sense of a self-fulfilling prophecy) that affect students' subjective expected probability of educational success (cf. papers 3 and 4). In case of reference-group effects on teachers' evaluations, supplemental *opportunity-mediated* mechanisms will come into play (cf. paper 2).

The remainder of this introduction will be structured as follows:

In section 2, I will begin with Max Weber's well-known definition of sociology in order to use it for a more profound elaboration on the debate on the consecutive prevalence of either *Erklären* (explanation) or *Verstehen* (understanding) as the methodological principle in the field which is denoted as *humanities* today. After that, I will outline how this debate connects to the concept of *social mechanisms* (Elster, 1985, 1989; Hedström and Swedberg, 1996, 1998a; Hedström, 2005; Hedström and Ylikoski, 2010, but also see Opp, 2005a) that tries to bridge the gap between *Erklären* and *Verstehen* by the use of *middle-range theories* (Merton, 1957; also see Boudon, 1991). It is shown that mechanism-based explanations do not fall behind the conceptual rigor of deductive-nomological explanations (Hempel, 1942; Hempel and Oppenheim, 1948) – but are superior to them in understanding actors' particular *desires* and *beliefs* (Hedström, 2005) in the context of their situational *opportunities* (ibid., also see Coleman, 1990, ch. 1; Esser, 1993a, ch. 6).

In section 3, I will use another statement by Max Weber about individuals' *life chances* as a starting point for first sketching the enduring debate about the demand for *meritocracy* in the educational system; and second, to relate that discourse to findings about *inequalities in educational opportunities* and the social mechanisms behind them. As I will show, the latter elaborations cast a shadow on theoretical accounts that overhasty defend the prevalence of 'merit' (in terms of achievement) over individuals' social backgrounds in explaining their actual life chances.

In section 4, I will first provide a brief summary of the four papers at hand; and second, I will reconstruct the underlying social mechanisms in all four studies while also referring to the implied action-theoretical assumptions. Finally, the conclusion in section 5 will offer an outlook for all four papers in particular and for social sciences theory in general.

2 Historical and Analytical Foundations of Social Action Theory

At the beginning of his seminal monograph *Economy and Society*, Max Weber defines sociology as follows:

> Sociology (in the sense in which this highly ambiguous word is used here) is a science which attempts the interpretive understanding of social action in order thereby to arrive at a causal explanation of its course and effects (Weber, 1964, p. 88).

What is translated here as "interpretive understanding" and "causal explanation" is what reads "deutend verstehen" and "ursächlich erklären" in the German original.[1]

By referring to both explanation and understanding as two equally important aims of sociology, Weber bridges an important gap between two schools of thought that have been concurrent to each other since the 19^{th} century. In the next subsection, I aim to sketch the historical conditions that set the stage for Max Weber's definition of sociology (and that might also account for the occasionally huge gap between quantitative and qualitative methodology in the social sciences; e.g. Adorno et al., 1976).

2.1 Historical Developments: "Erklären" vs. "Verstehen"

As natural sciences matured to becoming the dominant scientific discipline in the 19^{th} century, the humanistic studies had to deal with the issue that now an empiristic paradigm in favor of practices such as experiment and observation became the gold standard of scientific methodology.[2] Roughly speaking, this development divided humanities into two camps (von Wright, 1971, p. 3ff.): one holding the view that the

[1] "Soziologie soll heißen: eine Wissenschaft, welche soziales Handeln deutend verstehen und dadurch in seinem Ablauf und seinen Wirkungen ursächlich erklären will" (Weber, 1985, p. 1). Where possible, I will try to cite available English editions of the German classics, but in some cases – such as the one here –, additional quotations in German are inevitable, or an English edition was not available.

[2] "Positivism has gradually taken possession of the preliminary sciences of Physics and Biology, and in these the old system no longer prevails" (Comte, 1865, p. 12).

methodological standard of exact natural sciences should also be applied to humanities – an approach referred to as *positivism*, – and another, genuinely anti-positivist stand that rejects positivists' methodological monism and advocates a contrast between natural sciences and disciplines such as history (for which many of the following arguments were developed; see Dilthey, 1927) that aim "to grasp the individual and unique features of their objects" (von Wright, 1971, p. 5) – usually denoted as *hermeneutics*.

Historically, sociology was founded as a positivist discipline. Comte (1865) summarizes the intention of positivist sociology as follows: "The primary object, then, of positivism is twofold: to generalize our scientific conceptions, and to systematize the art of social life" (p. 3). The latter can be understood as consisting of "Thoughts, Feelings, and Action" (p. 8)[3], and in his emphasis on invariable laws that he assumes to underlie human action (p. 10), Comte already anticipates the deductive-nomological paradigm as it will be later set up by Hempel (1942) and Hempel and Oppenheim (1948).[4] Consequently, Comte denotes Sociology as a "physique sociale" (Comte, 1839, 46[e] Leçon) that should follow the positivist principles developed by the natural sciences.[5]

However, as a reaction to the methodological adoption of positivism, the hermeneutic approach engaged in postulating a methodological uniqueness of the humanities. As both von Wright (1971) and Apel (1979) note, Droysen (1857, p. 11) appears to be the first scholar using the dichotomy of *Erklären* vs. *Verstehen*: "Nach den Objekten und nach der Natur des menschlichen Denkens sind die drei möglichen wissenschaftlichen Methoden: die (philosophisch oder theologisch) spekulative, die mathematisch-physikalische, die historische. Ihr Wesen ist: zu *erkennen*, zu *erklären*, zu *verstehen*" (emphasis added).

While the methodological *trias* proposed by Droysen is less known today, Dilthey's dichotomy separating the *Geisteswissenschaften* from the natural sciences became more prominent. It has been argued elsewhere (von Wright, 1971, p. 173; Apel, 1979, p. 17) that the term *Geisteswissenschaften* first – i.e. in the plural form – appeared in Jacob H. W. Schiel's translation of John Stuart Mill's term "moral sciences" in his *System of Logic* (Mill, 1843, 1863). Dilthey might have adopted it from this monograph, but also Hume (1913) has to be named in this context: Actually, he already uses the terms "moral philosophy", "the science of human nature", and "the moral sciences" (Hume, 1913, section I; section VII, part I) to refer to what is called *Geisteswissenschaften* today.[6] Interestingly, and although Hume considers mathematics to be far more clear and determinate than the more ambigue moral sciences, he finally arrives at the conclusion that "their advantages and disadvantages nearly compensate each other" (ibid). This, of course, is an elemen-

[3] At this point, one could argue that Comte's emphasis on *thoughts* and *feelings* also anticipates the first two concepts of the *D*esires, *B*elifes, and *O*pportunities (DBO) action model by Hedström (2005) that will be described in more detail in section 2.2.

[4] "The importance that we attach to theories which teach the laws of phenomena, and give us the power of prevision, is chiefly due to the fact that they alone can regulate our otherwise blind action upon the world" (Comte, 1865, p. 11). I will come back to the particular epistemological importance of the term 'blind' in footnote 12.

[5] After Quetelet (1835, 21ff.) used the term "physique sociale" to denote the statistical analysis of social phenomena, Comte switched to the term "sociologie" (e.g. Comte, 1865, p. 27).

[6] Ayer (1952) notes that "the best part of John Stuart Mill's work consists in a development of the analyses carried out by Hume" (p. 55).

tary prerequisite for the Kantian unification of empiricism and rationalism[7], and it also anticipates Weber's later unification of the *Erklären* and *Verstehen* camps. Before, however, beginning with Dilthey's *Einleitung in die Geisteswissenschaften* (Dilthey, 1883, 1984), the triumphant procession of the latter term as a self-reference of the German humanities was unstoppable.

Whereas the early Dilthey (1880 and earlier; see Dilthey, 1984) did not make use of Droysen's (1957) distinction between explanation and understanding when already arguing against Comte's positivism, he later sharpens it even more: "We explain nature, but we understand psychic life" (as cited in Makreel, 1992, p. 134; original (German) citation in Dilthey, 1924, p. 144), and "We explain through purely intellectual processes, but we understand through the cooperation of all our psychic powers" (as cited in Makreel, 1992, p. 134; original (German) citation in Dilthey, 1924, p. 172).[8] With this distinction, Dilthey initiates a position that became dominant for the school of Neo-Kantianism in the late 19^{th}/early 20^{th} century.[9]

Regarded in this context, it is indeed notable that Weber's definition of sociology re-integrates both methodological paradigms again. In doing so, it comes close to the recent line of arguing of analytical sociology in terms of social mechanisms (see next subsection). Moreover, becoming aware of the fact that the state-of-the-art methodology for the next decade established to be one of (quite mechanistic; cf. Esser, 1996b; Machamer et al., 2000) universal laws (Hempel, 1942), it is even astonishing (see Apel, 1979, p. 40 for a similar line of arguing).

In his essay *On some categories of interpretative sociology*, Weber (1922, 1981) specifies the demand for a synthesis of explanation and understanding more concretely. One the one hand, Weber clearly argues in favor of rationality as an interpretative *and by that means* explanatory principle – being the response of sociology in a more and more rationalized world.[10] But on the other hand, Weber explicitly stresses that "the 'understanding' (*Verstehen*) of the context must always be verified, as far as possible, with the usual methods of causal attribution, before any interpretation, however plausible, becomes a valid 'intelligible explanation'" (Weber, 1981, p. 151). It is characteristic of such an 'intelligible explanation' that by relying on the principle of instrumental rationality, even 'irrational' processes such as stock market panics can be explained adequately: In that case, the rational ideal type of action would serve as a benchmark

[7]In his *Prolegomena to any future metaphysics*, Kant acknowledges Hume for having interrupted his own "dogmatic slumber" (Kant, 1902, Introduction).

[8]As Giuliani (2003, p. 10) has observed, Apel (1979, p. 18) erroneously attributes the first of the two above-quoted statements by Dilthey to his *Einleitung in die Geisteswissenschaften* (Dilthey, 1883) – where, however, the latter author does not elaborate on the dichotomy between explanation and understanding.

[9]The Neo-Kantians such as Rickert (1899, 1902) or Windelband (1894) on the one hand built on Dilthey's distinction between explanation and understanding, but on the other hand, they rejected his *psychologism*, as they called it (Apel, 1979, p. 36). For a further discussion of Dilthey's relation to the Neo-Kantian tradition see Jalbert (2008) and Šuber (2010).

[10]For a discussion of the notion of *rationalization* see Weber's chapter on *bureaucracy* in his *Economy and Society* (Weber, 1978, ch. XI.)

to determine what *would* have happened, *had* actors behaved rationally (p. 154).[11] It becomes evident that for Weber, both concepts appear to be intertwined: "Sociology must reject the assumption that 'understanding' (*Verstehen*) and causal 'explanation' have *no* relationship to one another" (p. 157). When Bourdieu (1988, p. 774f.) later writes, "Theory without empirical research is empty, empirical research without theory is blind"[12], then he advances an integrative view that suits both the Weberian definition of sociology and the mechanism-based stream of research in analytical sociology. But before I come to Hedström and his (and others') *plea for* (social) *mechanisms* (Elster, 1998), I first aim to outline the still dominant (e.g. Opp, 2005b) theoretical framework against which Hedström and his coevals are arguing.

In 1942, Hempel published his seminal paper *The function of general laws in history* (Hempel, 1942). The main "punchline" of this article is the thesis that what has later been called *deductive-nomological explanations* have to serve as a methodological principle for both the natural and the social sciences. His definition of such a general law reads as follows:

> "By a general law, we shall here understand a statement of universal conditional form which is capable of being confirmed or disconfirmed by suitable empirical findings. [...] In every case where an event of a specified kind C occurs at a certain place and time, an event of a specified kind E will occur at a place and time which is related in a specified manner to the place and time of the occurrence of the first event" (Hempel, 1942, p. 35).

Practically, that means that each explanation consists of a phenomenon to be explained – the *explanandum* –, a general law from which the conclusion can be derived, and the actual initial conditions. Opp (2005a, p. 174) provides the following illustrative example:

> *Law*: If political discontent and perceived personal influence are relatively intense, the frequency of participation in demonstrations is high.
> *Initial conditions*: In October 1989 discontent and perceived influence of the population of Leipzig increased.

> *Explanandum*: The participation in the demonstrations in Leipzig increased in October 1989.

In deterministic explanations, the event will always occur once the initial conditions are met and the general law holds. But in contrast to the natural sciences, such a view

[11] I will come back to the prospects of *counterfactual* explanations in social sciences in the conclusion section of this introduction (section 5).

[12] This, of course, is borrowed by Kant's famous unification of *empiricism* and *rationalism*: "Thoughts without content are empty, intuitions without concepts are blind. It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts" (Kant, 1850, p. 46). Note that already Hume (1913) used similar epistemological metaphors – while presumably both authors are influenced by Plato's *Allegory of the Cave*.

evidently is too restrictive for the social sciences where general laws are usually not found. Consequently, Hempel (1942) relaxes the assumption of general laws towards the type of *probability hypotheses* (p. 41) – meaning that the prediction of an event can be asserted only with a high probability.

Nonetheless, and in spite of this limitation, Hempel sticks to the claim that "general laws have quite analogous functions in history and in the natural sciences" (Hempel, 1942, p. 35), and that "history can 'grasp the unique individuality' of its objects of study no more and no less than can physics or chemistry" (p. 37, orig. emph.).

With this view, Hempel evidently opposes the paradigm of hermeneutics, but also falls somewhat back behind Max Weber's early synthesis of an *understanding explanation*[13] – especially when claiming together with his co-author, Paul Oppenheim:

> "But the existence of empathy on the part of the scientist is neither a *necessary* nor a *sufficient* condition for the *explanation, or the scientific understanding*, of any human action. It is *not necessary*, for the behavior of psychotics or of people belonging to a culture very different from that of the scientist may sometimes be explainable and predictable in terms of general principles even though the scientist who established or applied those principles may not be able to understand his subjects emphatically. And empathy is *not sufficient* to guarantee a sound explanation, for a strong feeling of empathy may exist even in cases where we completely misjudge a given personality" (Hempel and Oppenheim, 1948, p. 146, emph. added).

While for Weber, understanding is a necessary part of a scientific theory in order to reconstruct individual actors' motives, believes, and means (Balog, 2008, p. 79), this view is clearly rejected by Hempel. Unfortunately, this 'raw' version of the covering-law model still prevails in influential textbooks (e.g. Opp, 2005b).

The early Popper is not only more or less singing from the same hymn sheet as Hempel and Oppenheim – but even claims to be the originator of the above-described theory (Popper, 1945a, ch. 25, note 3). In his later writings, however, he only incidentally refers to the methodological unity of both human and natural sciences (cf. Böhm, 2008, p. 366; Riedel, 1978, p. 163). Even more, he appears to be quite skeptic about the reception of the deductive-nomological model (while still claiming ownership for it); and instead, he sets the ground for later sociological theory by introducing a *logic of the situation* that is connected to a *rationality principle* as the underlying theory of action:

> "A voluminous literature, which in my opinion has contributed little to the problem, has sprung from a mistaken criticism of my ideas on *historical explanation*. In section 12 of *Logik der Forschung* I discussed what I called 'causal explanation', or deductive explanation, a discussion which had been anticipated, without my being aware of it, by J.S. Mill, though perhaps a bit vaguely (because of his lack of distinction between an initial condition and a universal law) [...] I did not, however,

[13] Von Wright (1971, p. 7) describes Weber's approach in the following way: A "positivist coloring is combined with an emphasis on teleology ('*zweckrationales Handeln*') and empathic understanding ('*verstehende Soziologie*')".

> regard this particular analysis as especially important for *historical* explanation, and what I did regard as important needed some further years in which to mature. It was the problem of rationality (or the 'rationality principle' or the 'zero method' or the 'logic of the situation'). But for years the unimportant thesis – in a misinterpreted form – has, under the same name 'the deductive model', helped to generate a voluminous literature" (Popper, 1974, p. 117).[14]

The latter concepts that Popper here, interestingly, juxtaposes are already introduced in *The Open Society and its Enemies* wherein Popper (1945a, ch. 14) sketches a situational logic which actually comes close to approaches by Coleman (1990) and Esser (1993a) – models well-known in theoretical sociology today. Similarly to the Neo-Kantians such as Rickert or Windelband (though Popper would surely be very unhappy with that comparison), he argues against *psychologism* that he contrasts with a *logic of the situation* that is, in turn, "the method of economic analysis" (Popper, 1945a, p. 290). In his later writing *The Rationality Principle* (Popper, 1994 [first published in French 1967; also appeared as Popper, 1985]), the details of this approach are worked out more concretely. As Nadeau (1993, p. 450) points out, the *logic of the situation* is an explanatory scheme that is adequate for the social social sciences as is the deductive-nomological scheme for the natural sciences. Where in the latter field, a natural event is explained by the coincidence of an initial condition and a general law, in the social sciences, we have the explanandum of a social event-type that is explained by a scientific model or theory conditioning on the rationality principle. By the term 'zero method', Popper (1994) addresses that he does not claim that every actor at every time acts in a rational way – but that she acts in a manner which is adequate to the social situation as she herself sees it (which equals the *subjective* interpretation of the rationality principle; see Nadeau, 1993, p. 456). In *The Poverty of Historicism III*, Popper (1945b) writes:

> "I refer to the possibility of adopting, in the social sciences, what may be called the method of logical or rational construction, or perhaps the 'zero method'. By this I mean the method of constructing a model on the assumption of complete rationality (and perhaps also on the assumption of the complete possession of information) on the part of all individuals concerned, and of estimating the deviation of the actual behavior of people from the model behavior, using the latter as a kind of zero co-ordinate" (Popper, 1945b, p. 82).

Since "in most, if not all, social situations there is an element of *rationality*" (ibid.), it is possible to refer to a kind of *ideal type* of human behavior as the benchmark for sociological analysis. Note that this position remarkably resembles what already Weber (1922, 1981) has put forward in his *Gesammelte Aufsätze zur Wissenschaftslehre* wherein he introduces the concept of an "objectively correct rationality [*Richtigkeitsrationalität*]" (Weber, 1981, p. 154) as an instrument in order to draw inferences about human behavior. As mentioned, it distinguishes Weber from Hempel that the former refers to the

[14] Note, however, that already in the third part of his *Poverty of Historicism* (Popper, 1944a,b, 1945a), he wrote: "I do not intend to assert that there are no differences whatever between the methods of the theoretical natural and the social sciences; such differences clearly exist" (Popper, 1945b, p. 78).

principle of rationality as an instrument to *understand* human action; and when Popper incorporates this approach, he approximates the methodology of social sciences again to this endeavor.

Koertge (1979) proposed how Popper's situational logic can be quasi-formalized in analogy to Hempel's covering-law principle. While the latter takes the form (Hempel, 1965, p. 471)

> "A was in a situation of type C.
> A was a rational agent.
> In a situation of type C, any rational agent will do x.
> ___
> Therefore, A did x",

for Popper, she would propose something like

1. *Description of the situation*: Agent A was in a situation of type C.

2. *Analysis of the situation*: In a situation of type C, the appropriate thing to do is x.

3. *Rationality principle*: Agents always act appropriately to their situations.

4. *Explanandum*: (Therefore) A did x (Koertge, 1979, p. 87).

Comparing both statements, a striking difference is immediately evident: While Hempel (1965) speaks of a A as a rational agent, in the Popperian re-formulation as formalized by Koertge (1979), such a strong notion of rationality is replaced with A acting "appropriately" to her situation. This is not only coming close to Weber's notion of *Richtigkeitsrationalität* again, but due to the broader concept of rationality, it is also in accordance with the well-known *Thomas Theorem* reading "If men define situations as real, they are real in their consequences" (Thomas and Thomas, 1928, p. 572) – meaning that it is the actor's *subjective definition* of the situation that will create social reality.[15] Note that Popper himself doesn't become tired of highlighting the demand for an *understanding* of the actors' motives and beliefs – perhaps most pointed in his rationality essay: "The fundamental problem of both the theoretical and the historical social sciences is *to explain and understand events in terms of human actions and social situations*. The key term here is '*social situation*'" (Popper, 1994, p. 166; orig. emph.). Since by this line of reasoning, the methodological principle of social sciences is linked to the *Verstehen* tradition again (also see Hedström and Swedberg, 1998a, p. 350), Böhm (2008, p. 384) explicitly draws a parallel between Popper's situational logic and 18th century hermeneutics. Thus, on the one hand, in Popper's situational logic, not all people are *equally* rational in terms of a simple behaviorist input-output machinery given a

[15] I will come back to the *Thomas-Theorem* when discussing the social mechanisms to be unveiled in a self-fulfilling prophecy explanation.

particular social situation.[16] On the other hand, however, the method of interpretative sociology should not be simplified to psychologism; this is what would stand in conflict with Popper's demand for an objective situational logic coinciding with the method of economics (Hedström and Swedberg, 1998a, p. 347).

Because of the conceptual wideness of his approach, Hedström and Swedberg (1998a, p. 340) note that Popper "argued for a rationality-based analysis long before scholars such as James Coleman and Gary Becker, who usually are considered to be the intellectual forefathers of contemporary rational choice sociology".[17] However, it should not be ignored that also several authors objected against several statements in Popper's situation logic. For instance, Nadeau (1993) criticized that the status of the rationality principle which Popper (1994, p. 169) denotes as being both "almost empty" as well as "actually false, though a good approximation to truth" (p. 177) does not fulfill the requirements of *falsificationism* as demanded in his *Logic of Scientific Discovery* (Popper, 1959 [1935]).[18] Furthermore, Hedström et al. (1998, p. 354) note that Popper's social situations appears to be limited to one actor who finds herself confronted with some 'obstacles' she has to cope with. That is, Popper's social situation does neither really cover social interaction – nor does it consider actors' *interests* (Hedström et al., 1998, p. 354f.).

In order to solve these problems, Hedström et al. (1998, p. 357) propose an extension of Koertge's (1979) re-formalization of Popper's situational logic:

1. *Description of the situation*: Agent A was in a situation of type C characterized by a specific array of action alternatives, x_1 to x_n.

2. *Description of interests*: Agent A wants to attain end E.

[16] What can be noted for rational action theory in particular also holds for causal mechanisms in general: "[O]ne should not think of mechanisms as exclusively mechanical (push-pull) systems" (Machamer et al., 2000, p. 2).

[17] As Hedström and Swedberg (1998a, p. 340) point out, the fact that Popper's thoughts on situational logic and rationality were more or less ignored by sociologists can be explained by the hostility of sociology towards the notion of rationality during the 1970s and 1980s. A prominent example might be the following statement from the *Dialectic of Enlightenment*: "With the spread of the bourgeois commodity economy the dark horizon of myth is illuminated by the sun of calculating reason, beneath whose icy rays the seeds of the new barbarism are germinating" (Horkheimer and Adorno, 2002, p. 25). Critics of *Critical Theory* might counter with Mario Bunge's *bon mot*: "[W]hy is academia destructing itself by producing and diffusing 'postmodern' gobbledygook?" (Bunge, 1997, p. 413).

[18] The argument against Popper's rationality principle appears to end up in what is known as the *Münchhausen Trilemma*, i.e. the inescapability between a *circular argument*, an *infinite regress*, and an *axiomatic argument* (cf. Albert, 1985, ch. 1). Despite this, I hold the thesis that for Popper's notion of rationality as a principle that is "almost empty" (Popper, 1994, p. 169), a status applies similar to what Ingeborg Maus denoted as "the presuppositionlessness of modernity" when defending John Rawls' (equally criticized) concept of the *original position* in his *Theory of Justice* (Rawls, 1971): "Die 'gegenseitige Stützung vieler Erwägungen' nähert sich dem Grundmuster moderner Begründungen an: diese sind notwendig zirkulär, ohne dass es ihnen erlaubt wäre, in einem tautologischen Sinn selbstreferentiell zu sein" (Maus, 2006, p. 86; orig. emph.). Unfortunately, space constraints prevent me to elaborate more intensely on this issue.

3. *Description of beliefs*: Agent A has reasons to believe that action x_1 is the best way to attain E in situation C.

4. *Rationality principle*: Agents always act rationally; that is, they choose the course of action that they believe to be the best way of realizing their interests.

5. *Explanandum*: (Therefore) A did x_1.

Note that this enumeration is not only an extension of Popper's original conception but already moving a good deal away from it. For instance, apart from considering actors' interests and beliefs, the form of the rationality principle formulated here has not much to do anymore with Popper's "empty principle" which is "actually false, though a good approximation to truth". While Popper did not claim that every actor always acts rational, according to the fourth principle in the notation by Hedström et al. (1998), this is evidently the case (though the *type* of rationality is one of the weak sort).

Summing up their intention, Hedström et al. (1998) attempt to overcome some shortcomings of Popper's situational logic by providing an extended formalization also considering actors' interests and beliefs. However, the reader should be aware of the fact that the authors tend to use Popper's situational logic merely as a 'springboard' for their own action theory consisting of *D*esires, *B*eliefs and *O*pportunities (cf. Hedström, 2005, ch. 3; also see section 2.2 below). Having arrived at a suitable working definition of social mechanisms, this very basic action model will indeed prove useful in illustrating various types of social mechanisms (cf. section 2.2). Luckily, there are theorists who note that actors' beliefs and motives are implicitly part of the Popperian social situation (Nadeau, 1993; Böhm, 2008). Therefore, we are not entirely on the wrong track in following Hedström et al. (1998) in the above step, but particularly regarding the upcoming discussion of different notions of rational choice theory, we should keep in mind that Popper (1994) himself might not subscribe to every item in the enumeration above. Nonetheless, building on the *extended* deductive-nomological framework on the one hand ensures the capability of sociology of being both an understanding and an explaining academic discipline, and on the other hand, it is a valuable starting point for the following subsection on analytical sociology.

Interim conclusion This subsection has shown how Max Weber's (1964/1985) definition of sociology is deep-rooted in the German *Erklären-Verstehen* debate of the 19^{th} century. While the deductive-nomological explanation (or covering-law model) by Hempel and Oppenheim (1948) falls too short for an entire understanding of human action in the Weberian sense, Popper's *situational logic* (Popper, 1944a,b, 1945b) together with his rationality principle (Popper, 1994) provides a good starting point for an analytical theory of action that lives up to the demand for an *understanding* explanation by pointing to more fine-grained social mechanisms (Hedström et al., 1998; Hedström, 2005).

2.2 Analytical Sociology and Social Mechanisms

Although Popper's situational logic already sets the ground for later rational-action based reconstructions of individuals' social situation, for a long time after Popper, a whole stream of social science tended to reduce causal explanations on detecting significant correlations without providing a sufficient understanding explanation for them.

To given an illustrative example, Popper reports that when another distinguished scholar (of whom he does not tell us the name) once uttered at a scientific conference that science was just measuring and correlating results, he himself replied: "I suggested we should ask for a grant for a project of measuring the length, width, thickness, and weight of the books in the British Museum – in order to study possible correlations between these measurements. I predicted that we should be able to find strong positive correlations between the product of the first three measurements and the fourth" (Popper, 1994, p. 155).

As pointed out by Hedström and Swedberg (1996), a good starting point for the relevance of this kind of argument for the social sciences is provided by the controversy emerging from Boudon's monograph *Education, Opportunity, and Social Inequality* (Boudon, 1974), its review by Hauser (1976), and a rejoinder by Boudon (1976) again.

Hauser (1976) particularly criticizes Boudon's (1974) distinction between statistical and theoretical models and the idea that the latter should be used in order to explain what has been computed in the former. Boudon, however, replies that "we must go beyond the statistical relationships to explore the generative mechanisms responsible for them. This direction has a name: theory. And a goal: understanding" (Boudon, 1976, p. 1187). In a later writing, he adds that "causal analysis does not explain the chart. It simply summarizes it [...] *Understanding* a statistical structure means in many cases building a generating theory or model [...] that includes the observed empirical structure as one of its consequences" (Boudon, 1979, p. 51f., orig. emph.; also see Hedström and Swedberg, 1996, p. 292). As the observant reader might have noticed, the notion of *Verstehen* is explicitly brought up again in order to overcome a crucial shortcoming of merely correlational statistical analysis. Referring to von Wright (1971), Hedström and Swedberg (1996) hold the view that especially covering-law explanations are nothing more than "black-box explanations" (p. 297); and one may (and at least I do) regard the combination of a reductionist covering-law approach followed by short-sighted statistical analysis (i.e. "variable sociology"; also see Esser, 1996b) to be a particularly unholy alliance.

Consequently, Hedström (2005, p. 16) follows early critics of covering-law type explanations (such as Salmon, 1971) when conceding that the former neither go very far, nor are they generally considered to be acceptable scientific explanations. Thus, let us ask with Esser (1996b): *"What's wrong with variable sociology?"* As Goldthorpe (2001) points out, it has something to do with the implied notion of causality. Goldthorpe distinguishes three traditions of causal modeling in the social sciences: *robust dependence*, *consequential manipulation*, and the *generative process* account. The 'robust dependence' tradition became most prominent with the Wisconsin model of status attainment process (Blau and Duncan, 1967) wherein social stratification was modeled as a complex path structure focusing on the statistical correlations between the variables in the model. The 'consequential manipulation' tradition can be related to the Holland-Rosenbaum-Rubin model of a methodological[19] *counterfactualist treatment-effect approach* that evaluates the efficacy of a treatment – e.g. a job-training program – against an artificial situation wherein individuals had not received the treatment (see Gangl, 2010, for an excellent review of both the seminal econometric papers and numerous applications of the relevant methods in both economics and the social sciences). Both traditions may be subsumed beneath what Esser (1996b) called *variable sociology* – because in both cases, statistical techniques are regarded to be a sufficient approximation of (if not identical to) the implied idea of causality. To be precise, both traditions are *not explanatory* since the necessary explanatory link between the independent and the dependent variables is lacking; they are *incomplete* since the often-applied strategy of adding more covariates to overcome conceptual shortage will by no means ever be exhaustive; and they are *meaningless* since in most cases, a "general theory of decision making between given situational alternatives" (Esser, 1996b, p. 163) – entailing additional covariates that may mediate individual decision-making – is missing.[20] The third tradition, and lucky for us, the tradition that Goldthorpe (2001) regards to be suitable to overcome the above-described shortcomings is the *generative process* tradition. The latter, *nota bene*, is equivalent to the idea that the association between two variables X and Y is created by some 'mechanism' – i.e. that the concept of causation is tied "to some process existing in time and space, even if not perhaps directly observable, that actually generates the causal effect of X on Y and, in so doing, produces the statistical relationship that is empirically in evidence" (Goldthorpe, 2001, p. 9).

Goldthorpe argues that a sociological explanation standing in accordance with the generative process model should proceed along a three-phase sequence (Goldthorpe, 2001, p. 10):

1. establishing the phenomena that form the *explananda*

2. hypothesizing generative processes at the level of social action

3. testing the hypotheses.

In the first of the three phases (which Goldthorpe borrows from Merton, 1987), the researcher should unveil the social regularities that she aims to explain (e.g. variation in individuals' educational transition probabilities). Critique of a naïve use of statistical causal-modeling techniques should not be equated with a prohibition of inductive and explorative techniques such as scaling or clustering in order to establish *what* should be explained and tested thereafter.

[19]I denote this approach 'methodological counterfactualism' because it should be distinguished from an *ontological* counterfactualism as prominently held by Lewis (1973, 1977, 1979, 1981, also see section 5).

[20]Note that also Esser uses the term *verständlich* in this context (Esser, 1996b, p. 163) – unfortunately without further elaboration.

In the second phase, the causes for the social regularities have to be unfolded. Consistent with the paradigm of *methodological individualism*[21] before any statistical causal modeling, actors' reasons within specific social situations have to be captured. Anticipating what will be discussed more precisely in the next paragraph, this step can also be described with Elster's (1985, 1989) metaphor of *opening the black box and showing the cogs and wheels of the machinery.* Thus, the second phase is closely related to mechanism-based explanations.

Finally, the third phase involves the actual test of what has been merely hypothesized in the second phase. Importantly, this stage is subordinated to the preferably fine-grained theoretical explanation since it is conceptualized as an empirical examination of the former stage. Sophisticated statistical techniques as applied in the consequential-manipulation framework are admitted in this stage, but not as attempts to derive causal relations directly from data analysis (Goldthorpe, 2001, p. 11).

The focus of this introduction will lay on the second phase described by Goldthorpe (2001) in terms of unveiling potential causes for both regularities already observed in preceding studies (e.g. reference-group and teacher expectancy effects) and new hypotheses that were deduced from the phenomena already established. To a certain extent, this endeavor is in line with Coleman's (1990) metatheoretical plea for unveiling macro-micro hypotheses, action-theoretical assumptions on the micro-level, and micro-macro hypotheses in order to account for observed macro-level regularities (such as the Weberian 'Spirit of Capitalism'). However, the understanding of a social mechanism as defended here goes beyond both Coleman's (1990, p. 5) 'methodological pragmatism' and his action-theoretical restrictions (p. 18). Also to show this, a suitable working definition of a social mechanism will prove particularly helpful – which will be the focus of the next paragraph.

Definitions of social mechanisms While the term '*social mechanisms*' evidently refers to mechanism-based explanations in the social sciences, the more general notion *causal mechanism* reveals the crucial demand of this stream of research: getting closer to the actual causal structure than it is possible merely by means of correlational analysis.

Definitions of causal mechanisms have been proposed numerously, and therefore, reviews of these definitions are necessarily incomplete. Below, I supplement the summaries of the most important definitions provided by Gross (2009) as well as Hedström and Ylikoski (2010).[22]

[21] Yet, I tend to follow Hedström and Ylikoski (2010, p. 60) in linking social mechanisms to the slightly weaker concept of *structural individualism.* For a review of various approaches towards the notion of *individualism* see Udehn (2002).

[22] When Bunge (2004, p. 191) writes that "there are nearly as many systems theories as systems theorists", one could say the same about definitions of causal mechanisms – so the (supplemented) overview below is still not intended to be an exhaustive enumeration.

Table 1: Definitions of causal mechanisms

Author	Definition	Source
Bechtel & Abrahamsen	A mechanism is a structure performing a function by virtue of its component operations and their organization. The orchestrated functioning of the mechanism is responsible for one ore more phenomena.	Bechtel and Abrahamsen (2005); Bechtel (2006, 2008)
Bunge	A mechanism is a process in a concrete system that is capable of bringing about or preventing some change in the system.	Bunge (1997, 2004)
Glennan	A mechanism for a behavior is a complex system that produces that behavior by the interaction of several parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.	Glennan (2002)
Machamer, Darden and Craver	Mechanisms are entities and activities organized such that they produce regular changes from start to finish.	Machamer et al. (2000); Darden (2006); Craver (2007)
Elster I	A mechanism explains by opening the black box and showing the cogs and wheels of the internal machinery. A mechanism provides a continuous and contiguous chain of causal or intentional links between the *explanans* and the *explanandum*.	Elster (1985, 1989)
Elster II	Mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions.	Elster (1999)
Gambetta	Mechanisms have the form, 'Given certain conditions K, an agent will do x because of M with probability p.' M refers either to forms of reasoning governing decision making (of which rational choice models are a subset) or to subintentional processes that affect action both directly (as impulsiveness) or by shaping preferences or beliefs.	Gambetta (1998)
Gross	A social mechanism is a more or less general sequence or set of social events or processes analyzed at a lower order of complexity or aggregation by which – in certain circumstances – some cause X tends to bring about some effect Y in the realm of human social relations. This sequence or set may or may not be analytically reducible to the actions of individuals who enact it, may underwrite formal or substantive causal processes, and may be observed, unobserved, or in principle be unobservable.	Gross (2009)
Hedström	Mechanisms consist of entities (with their properties) and the activities that these entities engage in, either by themselves or in concert with other entities. These activities bring about change, and the type of change brought about depends on the properties of the entities and how the entities are organized spatially and temporally.	Hedström (2005)
Little	A causal mechanism is a series of events governed by lawlike regularities that lead from the *explanans* to the *explanandum*.	Little (1991)

continued

Table 1: Definitions of causal mechanisms

Author	**Definition**	**Source**
Opp	An *explanation by mechanisms* exists if it can be shown how a relationship between variables is brought about.	Opp (2005a)
Stinchcombe	Mechanism means (1) a piece of scientific reasoning which is independently verifiable and independently gives rise to theoretical reasoning, which (2) gives knowledge about a component process (generally one with units of analysis at a "lower level") of another theory (ordinarily a theory with units at a different "higher" level), thereby (3) increasing the suppleness, precision, complexity, elegance, or believability of the theory at the higher level without excessive "multiplication of entities" in that higher-level theory, (4) without doing too much violence (in the necessary simplification at the lower level to make the higher-level theory go) to what we know as the main facts at the lower level.	Stinchcombe (1991)
Woodward	A model for a mechanism (*a*) describes an organized or structured set of parts or components, where (*b*) the behavior of each component is described by a generalization that is invariant under interventions, and where (*c*) the generalizations governing each component are also independently changeable, and where (*d*) the representation allows us to see how, by virtue of (*a*), (*b*), (*c*), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves.	Woodward (2002)

Notes: Adopted from Hedström and Ylikoski (2010, p. 51). Definitions by Gross, Opp, and Stinchcombe added by myself.

Instead of going through each of these definitions separately, I rather prefer to review them systematically regarding their similarities and differences with respect to a set of fundamental dimensions. Before I do so, however, let us begin with Elster's pointed picture of "opening the black box and showing the "cogs and wheels of the internal machinery" as a starting point. The picture is accurate for the reason that it aptly describes what is actually lacking in variable sociology. It has something to do with what Merton (1957) called "sociological theories of the middle-range"; that is, developing more fine-grained explanations in order to theoretically account for the social explananda at hand (also see Boudon, 1991). It also appears to be sort of the least common denominator of the definitions at hand.

Going more into detail, conceptions of causal mechanisms appear to differ in at least 3 dimensions: i) *observability*, ii) *law-likeness*, and iii) *conceptual level of analysis*.

> i) *Observability*: While both Mahoney (2001) and Bunge (2004) hold the view that mechanisms refer to some kind of *unobservable* variables that account for the (observable) outcomes, Hedström and Ylikoski (2010) are critical towards such a restriction. A contrary position was taken by Reskin (2003) according to whom a mechanism is *always observable*. Though not

held overtly, this view is also implied in Opp (2005a) who unfortunately appears to be led a bit astray when he explicitly reduces social mechanisms on the quest for intervening variables (also see the critique in Hedström and Ylikoski, 2010, p. 51f.).[23] The view to be held in this introduction is to follow authors such as Gross (2009) and Hedström and Ylikoski (2010) in refraining from imposing any restrictions on the observability of social mechanisms. In methodological terms, while bringing back in individual-level explanations for simple macro-macro relations might be traceable by means of observable indicators, when looking for even more fine-grained explanations, we may sooner or later encounter a level whereon we are unable to operationalize our concepts within large-scale empirical studies. In epistemological terms, social mechanisms are of course intended to approximate the underlying causal structure by means of more fine-grained and partly observable entities – but as shown by Hume (1913) and Kant (1902), causality *itself* remains unobservable.

ii) *Law-likeness*: Opp (2005a) and Bunge (2004) very strictly connect social mechanisms to general laws: "No law, no possible mechanism; and no mechanism, no explanation" (Bunge, 2004, p. 207). Opp (2005a) opposes against Hedström and Swedberg's (1996, 1998b) critique of the covering-law model (Hempel, 1942; Hempel and Oppenheim, 1948; Hempel, 1965), and he suggests to "complement the H[empel]O[ppenheim]-scheme by a methodological postulate referring to explanations by mechanisms" (Opp, 2005a, p. 176; addenda in square brackets by myself) – which would, according to Opp (2005a), be superior to rejecting the whole scheme in lack of a true theoretical alternative (ibid.).

In contrast, apart from Hedström, also Elster (1998, p. 48) observes a difference between an explanation with laws and an explanation by mechanisms: "[A] law has the form 'If conditions C_1, C_2, ..., C_n obtain, then *always* E.' [...] [A] statement about mechanisms might be 'If C_1, C_2, ..., C_n obtain, then *sometimes* E (emphasis added by myself). As Opp (2005a, p. 177) correctly notes, this account of a social mechanism basically does not differ from a statistical, probabilistic or non-determinist law, so this appears to be the wrong track. The position taken here is that while the latter critique is obviously justified, the claim that mechanism-based explanations are still in line with the covering-law model since they can be regarded as a simple complement of them misses, in my view, the point. Relativity theory still relies, to a great deal, on Newtonian mechanics – but one would abstain from denying it the status of a unique theoretical approach. Likewise, mechanism-based explanations revealed an important shortcoming of the covering-law model, i.e. the

[23] As Esser (1996b, p. 160) points out, it is an attribute of variable sociology to add background variables to quantitative data analysis in an *ad hoc* and unsystematic manner. This of course falls back behind the demand of mechanism-based explanations aspiring to unveil the respective generative processes (Goldthorpe, 2001).

tendency to neglect what's happening beyond the surface. One could refer to the 'realist' approach in the philosophy of social sciences (Bhaskar, 1975; Collier, 1989, 1994) which most overtly holds the view that only by specification of mechanisms, scholars move from the *empirical* domain of correlation to the *actual* domain of the generative processes or events, and finally on to the *real* domain wherein the general *causal* mechanisms are located (Gross, 2009, p. 361; also see Kemp and Holmwood, 2003 and Maxwell, 2004). This is accomplished by no longer restricting explanations to correlations of single factors but by specifying the underlying *generative processes* which can be split up in smaller entities (Machamer et al., 2000; Mayntz, 2004).[24] In a suchlike understanding of social mechanisms, one would not be satisfied with Coleman's (1990) 'methodological pragmatism' stating that "[t]he criterion is instead pragmatic: The explanation is satisfactory if it is useful for the particular kinds of intervention for which it is intended" (Coleman, 1990, p. 5). Rather, the more fine-grained an explanation is split into smaller entities, the closer it is able to approximate towards causality. Hence, in the terminology of Lakatos (1978), mechanism-based explanations are a new *research program* tackling the protective belt of the reductionist covering-law model.[25]

iii) *Conceptual level of analysis*: Proponents of mechanism-based explanations who feel uncomfortable about the non-negligible coincidence of the former and rational-choice or subjective-expected-utility theories of action occasionally see the need for proposing a 'new' definition of social mechanisms that overcomes this restriction (e.g. Gross, 2009). However, such a conclusion is not necessarily true: For instance, Bunge (1997, 2004) proposed a conceptualization of social mechanisms that is in line with his more general account of *systemism* – which is not restricted to *methodological individualism*. Moreover, also Hedström and Ylikoski (2010) – themselves proposing the *D*esires, *B*eliefs, and *O*pportunities model (cf. below) to account for individual-level action – would not like to see mechanism-based explanations to be restricted to theoretical models of the latter kind. More specifically, Hedström and Ylikoski (2010) argue for what has been labeled *structural individualism* (Wippler, 1978): As opposed to methodological individualism, structural individualism explicitly takes potential situational constraints that

[24] One might object against the final step of realist philosophy of science that even when theorizing more and more fine-grained, 'real' causality will never be reached without divagating into metaphysics again (see e.g. Glennan, 1996, p. 65).

[25] Udehn (2002, p. 502) also notes that "the current emphasis on social mechanisms [...] may be seen as a sign of the decreasing importance attached to laws in social sciences, especially sociology". A Lakatosian account to mechanism-based explanations regarded from a standpoint of scientific evolution might be even more justified in the light of Kelle and Lüdemann's (1995) analogous approach towards the role of *bridge assumptions* in rational action theory. Rigid readers of Popper (1959) might furthermore argue that Stinchcombe (1991) is on the wrong track in his demand for *verification* of scientific reasoning.

> may be located on a higher level-of-analysis than the individual level into account (also see Lindenberg, 1990, p. 737f.). It is therefore compatible with the Popperian situational logic as well as with more elaborate micro-macro models proposed by Coleman (1990) and Esser (1993; also see Udehn, 2002).[26] Hence, although mechanism-based explanations can basically be related to different analytical approaches (Bunge, 1997, 2004), for the purpose at hand, the structural version of individualism is the adequate paradigm for explaining (and understanding!) both the causes and the effects of teachers' evaluations on different conceptual levels of analysis.

Altogether, I follow Gambetta's definition that social mechanisms have the form, 'Given certain conditions K, an agent will do x because of M with probability p' (Gambetta, 1998). I amend this definition by, *first*, assuming that M may either be observed or unobserved – which ensures that mechanism-based explanations are not reduced to *ad hoc* covariate controls. *Second*, although the probabilistic phrasing appears to point to the statistical interpretation of the covering-law model, it should be evident that by considering smaller entities of a *generative process* instead of correlations of factors, its distance to causality is lessened – which is why it should not be equated with the former.[27] Hence, the definition of social mechanisms defended here crucially distinguishes from the definitions proposed by Little (1991), Bunge (2004), and Opp (2005a) who stick to a 'law-like' character also of (social or causal) mechanisms. Also, although the meta-theoretical macro-micro-macro scheme proposed by Coleman (1990) definitely set the ground for the recurring plea for mechanism-based explanations, the latter go beyond Coleman's (1990, p. 5) 'methodological pragmatism' in aspiring towards more and more fine-grained explanations in order to get closer to the notion of causality. *Third*, I assume that mechanisms may refer to different levels of analysis or action theory paradigms. What should be avoided is to use mechanism definitions that are restricted to either individual- or contextual-only level of analysis. For instance, Bunge's (1997, 2004) system-related definition would not suit an action theory located on the individual level. As a consequence, a suchlike definition would necessarily have to remain silent on mechanisms due to actors' *desires* or *beliefs* (see below). On the other hand, as reference-group effects are one of the topics covered in the volume at hand, *structural individualism* (that is also open for context effects) is preferred to simple methodological individualism. Regarding action theory, it should be emphasized that the second part of Gambetta's (1998) definition is notably far-sighted regarding the distinction between more rational decision-making and the more subintentional processes that precede individual action as well. In section 4.1 of this introduction, I will build on the assumptions of the *M*odel of *F*rame *S*election (MFS; Esser, 1996b, 2010; Esser and Kroneberg, 2010; Kroneberg, 2006; Kroneberg et al., 2008, 2010; Kroneberg, 2011). This theory synthesizes arguments from the 'interpretative paradigm' about unconscious automatic processing in everyday life

[26] This will be of particular importance for also integrating *opportunity-mediated mechanisms* (see below; cf. Hedström, 2005, p. 55f.).

[27] Yet, a probabilistic notation of course facilitates implementation of testable hypotheses deduced from mechanism-based explanations.

(Esser, 1993b) and assumptions of rational-choice theory about actors' intentional and reflective cost-benefit weighting especially in high-cost situations. Hence, Gambetta's (1998) definition suits these propositions to be specified later on.

However, while at the beginning of his chapter, Gambetta (1998, p. 104) postulates that sticking to a rational-choice explanation as the most general mechanism, it will not only "be parsimonious and generalizable; it will also be the end of the story", after elaborating on the particular social mechanisms in the Italian academic system, individual decision mechanisms in education, and several alternative explanations regarding the relative deprivation theory suggested by Stouffer (1950), the author arrives at a more cautious conclusion: "[T]he findings suggest that something more than a rational mechanism is at work here and that the mechanisms [..], interesting as they may be in their own right, are not prima facie as parsimonious as the relative deprivation or emulation mechanisms" (Gambetta, 1998, p. 118).[28] Therefore, the above definition provides only the starting point for an approximation towards the core of a social mechanism – which might become more evident if particular *types* of social mechanisms according to different *notions* of rationality are inspected.

Action theory Hedström (2005) proposes the *D*esires, *B*eliefs and *O*pportunities (henceforth *DBO*) model as a very basic action theory in order to build up mechanism-based explanations of human action. Also for Hedström (2005), *actions* imply intentionality as distinguished from unintentional *behavior* – a dichotomy well-known from Weber. To put it brief, *beliefs* are propositions of the world held to be true, a *desire* can be described as a wish or want, and *opportunities* refer to the "'menu' of action alternatives available to the actor" (Hedström, 2005, p. 38f.). Important beliefs are those held about different alternatives of action at hand or about the probability of certain consequences that may emerge from different actions. Taken together, beliefs and desires are a "compelling reason" or have a "motivational force" (Hedström, 2005, p. 39). However, although the author lays emphasis on the fact that opportunities must always be known to the actors and thus influence actions via their beliefs (ibid.), desires and beliefs are not sufficient in explaining human action since opportunities exist independently of them. Hence, the *DBO*-model is a good example of a structural-individualist action theory that does not lose track of 'what's going on on the macro level'.

Since the *DBO* model assumes that individual action emerges in accordance with actors' desires and beliefs, given particular situational opportunities, we won't be completely mistaken in subsuming it under the concept of (weak) *rational* action models. More concretely, Hedström (2005) emphasizes that his model does not assume actors always to act *rationally*, but they're supposed to act *reasonably* and intentionally (Hed-

[28] The emulation mechanism is an alternative explanation for the fact that in the American Military Police at times of World War II, though occupational opportunities were much lower than in the Air Corps, soldiers were more satisfied with the promotion system there than in the latter. The emulation mechanism is comparable to what Beck (1986) labeled *Fahrstuhlhypothese* ['elevator effect'] in his *Risikogesellschaft* (Beck, 1986; Engl. *Risk Society*; Beck, 1992): that the subjective assessment that an educational achievement is essential is higher once the number of people in favor of that achievement is sufficiently high (also see Gambetta, 1998, p. 117).

ström, 2005, p. 60f.). This matches Boudon's critique of instrumental rationality and his plea for a model of *cognitive* or *axiological* rationality (Boudon, 1996, 1998, 2003). Concretely, *cognitive rationality* confers an observation that Boudon has noted for history of science in general on the fundamental prerequisites of action theory: that it has also to deal with evidently false beliefs of individuals[29]; and that this purpose can be resolved by assuming that the former can be reasonably reconstructed (Boudon, 2003, p. 12). Complementary to cognitivist rationality, *axiological* rationality can be understood in sense of Weber's *Wertrationalität* – which means, following Boudon (2003, p. 14), that "prescriptive *beliefs* are grounded in the mind of social actors on systems of reasons perceived by them as strong" (emph. added).

The line of attack of these pleadings are aimed at a too narrow notion of rationality that can be traced back to neoclassical economics. As Goldthorpe (1998, p. 169) nicely points out, rational action theories can be distinguished "according to whether they

(i) have strong rather than weak rationality requirements;

(ii) focus on situational rather than procedural rationality; and

(iii) claim to provide a general rather [than] a special theory of action."

First, while neoclassical economics holds a very strong notion of rationality assuming evidently unrealistic axioms such as individuals' perfect information or their strict maximization of utility (Becker, 1976), less restricted approaches lay emphasis on individuals' *bounded rationality* (Simon, 1955, 1957) and consequently demand from sociological explanations also to rationally reconstruct false beliefs (Boudon, 1996, 1998).

Second, in mainstream economics, human action is situationally constrained to such an extreme degree that in a given market situation, and given individuals' set of preferences, an actor will always maximize her utility – implying that choice is a result of rather automatic computation.[30] While also Popper's situational logic can be subsumed among the type of theories imposing strict external constraints on human action (though holding a much weaker concept of rationality), Simon (1955, 1957), Lindenberg (1985, 1990) and Lindenberg and Frey (1993) uncouple action from situational constraints by deducing the idea of subjective rationality from more psychological foundations. The same scheme applies to Boudon's cognitivist action theory (Boudon, 1996).

Third, some theories such as Gary Becker's "economic imperialism" (Goldthorpe, 1998, p. 175) claim that rational action theory (and, in Becker's version, even in a very strong and mechanistic form) suits for explaining various aspects of social life by means of consumption theory. In contrast, social science theorists like Coleman (1990) and Boudon (1996, 1998) are more skeptical in this regard – or, to phrase it differently, they are more aware of social science theories' explanatory limits.

---

[29] This will be of particular importance concerning a mechanism-based explanation of self-fulfilling prophecies in terms of a teacher treatment effect (cf. section 4.3).

[30] As Goldthorpe (1998, p. 175) observes in accordance to several other authors, "the paradox arises that the theory of 'rational choice' *par excellence* turns out to imply that little real choice in fact exists."

Goldthorpe (1998) himself pleads for a notion of subjective rationality of intermediate strength within strong bonds to situational restrictions. He regards rational action theory to be a *privileged* theory, "[...] that is, not just one theory of action among others but rather *the* theory with which attempts at explaining social action should start and with which they should remain for as long as possible" (Goldthorpe, 1998, p. 184). Notably, he does not become tired of stressing that exactly this is the demand for any *verstehende* sociology in the sense of Weber: By applying rational action theory of intermediate strength in order to rationally reconstruct actors' desires and beliefs within a given situational context, social scientists are neither led astray by simple input-output types of black box explanations that may empty into causally deficient variable sociology, nor are they confronted by too restricted psychologism or a too sloppy notion of rationality that may end up in the tautologism of approving every human behavior as rational *per se*.

Not to soften the core of rationality too much is also defended by Lindenberg (1990): According to his *method of decreasing abstraction*, theorists should actually start with a comparably tight notion of rationality that is subsequently amended by *bridge assumptions* in order to account for explanatory blind spots of simple rational choice theory models ("as simple as possible and as complex as necessary"; see Lindenberg, 1990, p. 738). To be precise, maintaining a relatively simple core of rational choice theory is desirable according to the premise of parsimoniousness well-known as *Occam's Razor* (Thorburn, 1918; Feuer, 1957; Popper, 1959, ch. 7) – but if the simplicity assumption is too unrealistic, the (empty) core of rationality can be cautiously widened by more realistic auxiliary assumptions. Since it was argued above that the model of man defended by what has been called "economic imperialism" is hardly suited even to come close to an *understanding* of individual action, a notable extension of the former model will be inevitable. Concretely, the notion of rationality defended here is one of the broader type that also strives to reconstruct actors' false beliefs by discarding untenable action-theoretic axioms (such as the postulate of individuals' perfect information); and it is further argued that this aim can be achieved best by reasonably reconstructing actors' beliefs in line with Weber's concept of *Wertrationalität*. In contrast, we should not loose sight of the premise of parsimony that will prevent us from becoming too sloppy in relaxing our rationality assumption.

Having argued for a rational action model that neither defends unrealistic assumptions nor stands at risk of getting cut down by Occam's razor, we have of course to ask how the preceding elaborations relate to the above-defined notion of social mechanisms. At the end of the last paragraph, we ended up in a discussion about whether the assumption of actors' rational choice would already be the end of a mechanism-based explanation. What we may conclude now is that at least a too tight notion of rationality will not suffice, but the precise form of rational-choice approach is still not sufficiently specified. The solution to this problem is that following Lindenberg's (1990) method of decreasing abstraction, we may introduce different auxiliary assumptions that approximate individual behavior depending on the type of social situation. For instance, if not only the subjective values of some decision alternatives at hand, but also their relative probabilities are known, maximizing their expected value by weighting the subjective value of each alternative by its probability may be a reasonable strategy. If, however, these probabilities are *not* known, several decision heuristics – potentially yielding quite different results – have been proposed (Coombs et al., 1970, ch. 5; Lave and March, 1975, pp. 140-143; Thorngate, 1980): There may be optimistic decision rules such as *maximax* (“Maximize the maximum possible value!”), pessimistic decision rules such as *maximin* (“Maximize the lowest possible value!”), minimum-variation decision rules such as *minirange* (“Minimize surprise!”), or minimum-regret decision rules such as *minimax regret* (“Minimize the maximum regret!”). All of them might be very rational in a particular social situation.[31] Hence, without additional auxiliary assumptions that amend the empty principle of rationality, mechanism-based explanations would stand at risk of being either theoretically underspecified and tautological, or unrealistic.

Getting back to Hedström’s (2005) very basic *DBO* model outlined above, the author lays emphasis on the fact that rational choice theory can be regarded as “a specific type of DBO theory” (Hedström, 2005, p. 41). Hence, I will first summarize the author’s presentation of different types of social mechanisms along the basic entities of DBO theory. Where required, however, I will then add the necessary bridge assumptions in order to reconstruct the implied social mechanisms in the four papers at hand (see section 4).

Types of social mechanisms While behavioral patterns (to be accounted for by sociological theory) can be split up in environmental effects, selection effects, and social interaction effects (Hedström, 2005, p. 47), cautious readers of Weber might think of his famous umbrella example[32] whereby he illustratively distinguishes uniform human behavior from uniform social action and therefore accompany the author in his conclusion that only the latter type of pattern will be the one to be accounted for by social mechanisms.[33] Within the pattern of social interaction effects, *desire-mediated*, *belief-mediated*, and *opportunity-mediated* interaction can be distinguished and form the dimensions wherein different types of social mechanisms can be classified.

1. *The* seminal case of a belief-mediated mechanism is Merton’s idea of a *self-fulfilling prophecy* (Merton, 1948) according to which initially false beliefs lead to an outcome that makes the false beliefs come true. In the example Merton (1948) uses, he asks the reader to think of a bank that is untruly said to be bankrupt. Once the rumor is in the world, the bank’s clients will all but hurry up to withdraw their savings – which might then unexpectedly lead to the bank’s actual insolvency.

[31] Below I will briefly discuss how rational-choice-based theories on social inequality in educational opportunities (Breen and Goldthorpe, 1997; Esser, 1999) relate to particular decision heuristics (cf. section 3).

[32] “Thus, if at a beginning of a shower a number of people on the street put up their umbrella at the same time, this would not ordinarily be a case of action mutually oriented to that of each other, but rather of all reacting in the same way to the like need of protection from the rain” (Weber, 1968, p. 23).

[33] I am not satisfied with Hedström’s (2005) conclusion to exclude selection effects from mechanism-based explanations. In the applied part of this introductory chapter (section 4), I will come back to a mechanism-based reconstruction of potential selection effects in two of the four papers at hand.

In course of the self-fulfilling prophecy, the bank crash might be amplified by an additional mechanism of *rational imitation*: The higher the number of others perceived by an individual who withdraw their savings, the higher the individual's own probability to do likewise (cf. Hedström, 2005, p. 48f.). It is evident that the mechanism of rational imitation strongly hinges on the distribution of individual thresholds (Granovetter, 1978; Granovetter and Soong, 1983) in that for a more 'jumpy' person, the observation of only a few others withdrawing their savings may suffice to make her join, while a more 'relaxed' coeval might only react if already a majority of persons participated in the bank run.[34]

Another very-well known example of a belief-mediated mechanism is Festinger's (1957) theory of *cognitive dissonance*.[35] The crucial idea behind that particular type of social mechanism is that if an individual is exposed to two cognitive elements opposing each other, a mental state of *cognitive dissonance occurs*. Since states of cognitive dissonance are costly in psychological terms, the individual strives to avoid them (also see Hedström, 2005, p. 51). For instance, a person holding a liberal political position might change her opinion (or at least modify it) if a *significant other* she interacts with, or the media she is exposed to, hold a conservative view – and reversely (see e.g. Beck, 1991; Mutz, 2002; Feldman and Price, 2008). Another option would be to persuade others of one's own opinion (Hedström, 2005, p. 51) – which might of course be more challenging depending on the number of people one interacts with.[36]

2. The second type of social mechanisms are *desire-mediated* mechanisms. Others doing **A** may increase my probability of doing **A** if i) their doing **A** influences how strongly I desire **A**; ii) I desire to be like (or unlike) them; or iii) I believe that doing the same as they do increases (or decreases) my chances of getting **B**, which I desire (Hedström, 2005, p. 52). While subtypes i) and ii) – differing from each other in that in i), actions are causes, and in ii), actions are objects of egos desires – the actor has a *primary* desire to act in accordance with others, in iii), the desire is of a *secondary* type (ibid). Examples for the different subtypes might be i) the choice of a mobile phone network contract given that same-network calls are cheaper than others; ii) adoption of fashion (or other) trends from celebrities; iii) joining a political party in order to increase one's chances to obtain a particular occupational position.

3. Finally, *opportunity-mediated* mechanisms occur if individuals' opportunity structure is the essential cause of their action. In contrast to simple environmental effects (such as rainfall), opportunity-mediated mechanisms are also the outcome

[34] Granovetter's threshold model is also a useful tool for specifying the mechanism of imitation regarding the emergence of social movements (Opp, 1991; Braun, 1995).

[35] See Heider (1946) for a similar idea published before Festinger (1957).

[36] Note that the mechanism of dissonance reduction may also be subsumed beneath the concept of a *desire*-mediated mechanism when *ego* does not adjust her *beliefs*, but her *wishes or wants* according to those of her significant others.

> of social interaction. For instance, considering the example of rational imitation given before, in the model of individual thresholds, the number of other people who have already acted (e.g. withdrawn their earnings, joined a social movement, etc.) are of course part of ego's opportunity structure – but further exogenous parameters may influence both ego's and alter's individual threshold (e.g. global financial crises, or a nation's level of repression).[37]

Figure 1 is taken from Hedström (2005, p. 59) and summarizes several types of social mechanisms. The upper half of the figure has a more illustrative intention in introducing different belief- and desire-mediated mechanisms for a single actor's states of mind. For instance, in case of *wishful thinking*, the actor tends to desire what she beliefs. A soccer fan desperately believing her team to win the next match would be an example. An example for adaptive desire formation is given by Elster (2007, p. 176): "If beautiful women reject my advances, I may console myself by the thought that by virtue of their narcissism they are actually the least desirable partners."

The lower half of figure 1 will provide more useful for the following reconstruction of the social mechanisms in the four papers of this volume (section 4) in that social mechanisms invoked by both actors' states of mind and opportunities during *social interaction* are addressed. For instance, the mechanism of *dissonance reduction* takes effect via alter's A_j influencing ego's wanting of A_i (while ego's beliefs remain constant). Regarding *rational imitation*, Hedström et al. (1998) provides the example of restaurant visitors using the fact whether a restaurant is crowded as an indicator of the menu quality – which is why crowded restaurants are desired more. And finally, the *self-fulfilling prophecy* is characterized by a concatenation of at least two distinct belief-mediated mechanisms with constant desires.

Interim conclusion Whereas in *variable sociology* (cf. Esser, 1996b), causation is confounded with significant correlations between variables, the *generative-process* tradition in sociology strives to approximate causation by uncovering some process that produces correlations between variables (Goldthorpe, 2001). Proponents of *mechanism-based explanations* refer to this procedure by invoking the image of "opening the black box and showing the cogs and wheels of the internal machinery" (Elster, 1985, p. 5). More specific definitions of social mechanisms differ according to the dimensions of i) observability; ii) law-likeness; and iii) conceptual level of analysis. Promising (though not necessary) action theories to be used for mechanism-based explanations are models of rational action theory which can be distinguished according to whether they i) have strong rather than weak rationality requirements; ii) focus on situational rather than procedural rationality; and iii) claim to provide a general rather than a special theory of action (Goldthorpe, 1998). Social mechanisms *as understood here* may consist of

[37] One example given by Hedström (2005, p. 55) concerns the explanation of differences in social mobility rates (between nations, or over time) by recurring to differences in mobility opportunities in terms of differing occupational or class distributions. I will come back to this issue in the next section.

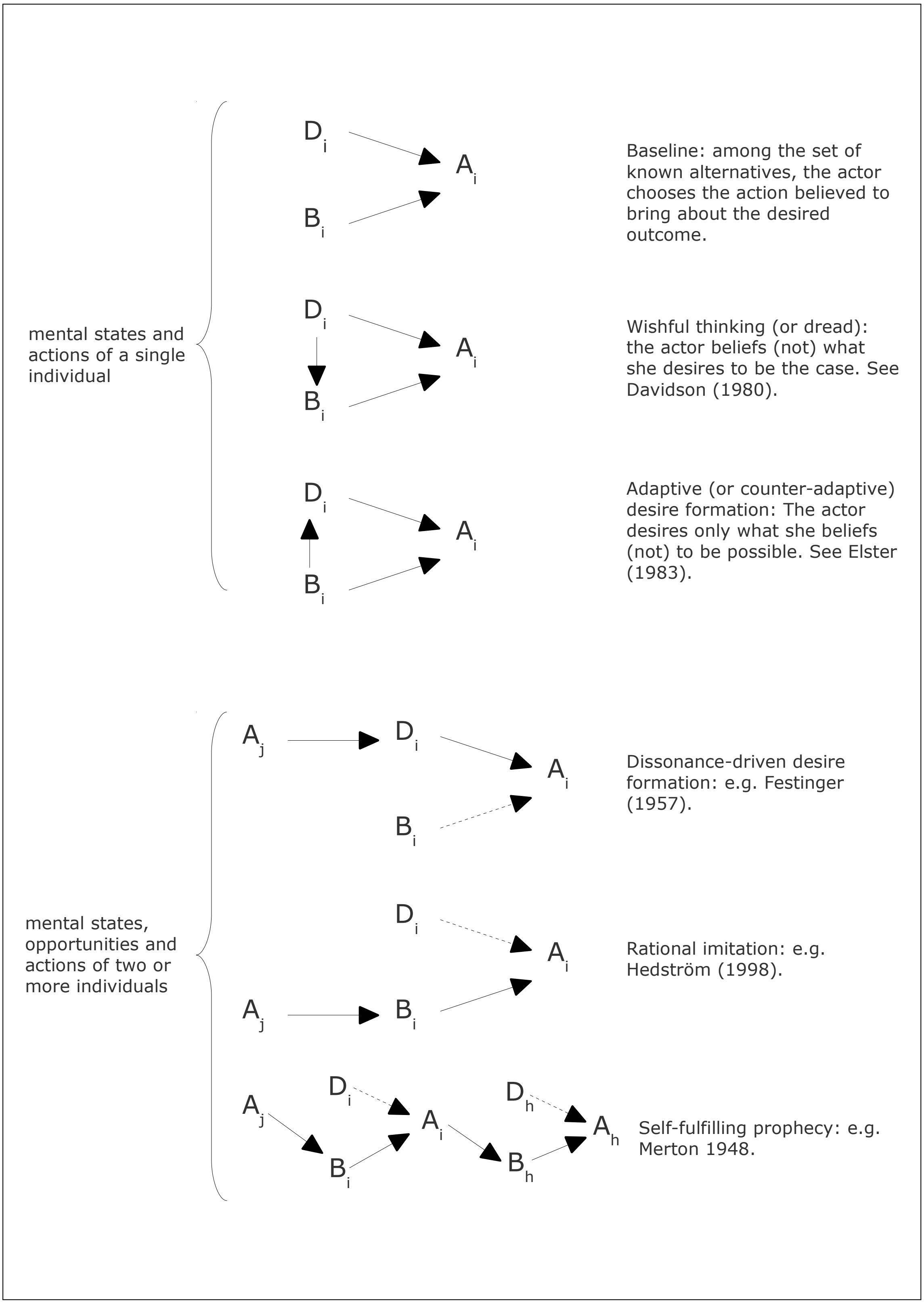


Figure 1: Types of social mechanisms. Taken from Hedström (2005, p. 59).

smaller entities (with decreasing chance of observability the more fine-grained the mechanism; see Machamer et al., 2000; Mayntz, 2004), should therefore not be equated with the covering-law model but fit within the model of *structural individualism* (Wippler, 1978; Lindenberg, 1990). The action theory of choice is one of intermediate strength with strong bonds to situational restrictions (Goldthorpe, 1998). The gold standard of *Occam's Razor* suggests to start with a simplified version that may be extended by bridge assumptions where found to be unrealistic (Lindenberg, 1990). A very basic *desires*, *beliefs* and *opportunities* model suffices to classify the most general types of social mechanisms such as *self-fulfilling prophecies*, *rational imitation* or *cognitive dissonance* (Hedström, 2005). Before belief- and opportunity-mediated mechanisms will prove useful for approximating the causes and effects of teachers' evaluations in the four papers of this volume, I will first summarize the discourse of whether educational systems are or should be meritocratic, and I will then contrast this discourse with contemporary theories of inequality in educational opportunities and their implied social mechanisms.

3 *Life Chances*, *Meritocracy*, and Inequality in Educational Opportunities

Having outlined the above foundations in the philosophy of science, in this section, I will begin with Max Weber's definition of social class as shaping individuals' *life chances*, and I will then discuss the notion of *meritocracy* that is still thought to be a justified mechanism for the allocation of societal positions. Quite contrarily, theories of social *i*nequalities in *e*ducational *o*pportunities (IEO) keep laying emphasis on the fact that theoretical accounts simply reducing status attainment processes on individuals' merit fall too short in neglecting considerable social background effects. This sets the ground for the subsequent reconstruction of the social mechanisms approximating the causes and effects of teachers' evaluations as one of students' *significant others* in their status attainment process (measured here in terms of educational transitions; see section 4).

Life chances Just as in the case of the theoretical foundations of social sciences in general, let us begin with a well-known quotation of Max Weber. His definition of social class reads as follows:

> "We may speak of a "class" when (1) a number of people have in common a specific causal component of their *life chance*, insofar as (2) this component is represented exclusively by economic interests in the possession of goods and opportunities for income, and (3) is represented under the conditions of the commodity of labor markets. This is "class situation". It is the most elemental economic fact that the way in which the disposition over material property is distributed among a plurality of people, meeting competitively in the market for the purpose of exchange, in itself creates specific *life chances*" (Weber, 2002, p. 33f., orig. quotations, my italics).

What becomes evident from this quotation is that Weber regards class situation as being determined by market situation; that is, people's position in the market itself (determined by their disposition over property) is the crucial factor for class positions with their respective life chances. From a methodological point of view, a pleasant implication of this classification is that the analysis of individuals' life chances can be accomplished by means of economic instruments such as rational action theory, which will be discussed below. From a normative point of view, the question that of course arises from Weber's definition is how society *should* equip individuals with particular life chances.

As Davis and Moore (1945, p. 242) have argued, it is inevitable for each society to deal with two problems: 1) Why do different positions carry different degrees of prestige, and 2) how do certain individuals get into these positions? Regarding the first question, the answer is that some positions are more agreeable, functionally more important and/or require more talent than others; and society keeps these positions attractive by different kinds of rewards (sustenance and comfort; humor and diversification; self-respect and ego expansion).

Regarding the second question, the answer to the first question already revealed that individuals' *talent* is an issue; and therefore, former ascriptive selection mechanisms are more and more replaced by others that are based on individuals *merits*:

Already in Plato's *Meno* (Plato, 2009) we come to know that even a slave who has not received any form of geometric education before can solve a relatively complex geometric task[38], and also in his *Republic*, Plato (1991) shows a comparable line of arguing in that each social position should be filled by talent and not by social origin (Riesman, 1967).[39] Similarly, also Weber (1968, p. 241) observes (also see Becker and Hadjar, 2011, p. 55f.):

> Today, the certificate of education becomes what the test for ancestors has been in the past, at least where the nobility has remained powerful: a prerequisite for equality of birth, a qualification for a canonship, and for state office. The development of the diploma from universities, and business and engineering colleges, and the universal clamor for the creation of educational certificates in all fields make for the formation of a privileged stratum in bureaus and in offices (...) When we hear from all sides the demand for an introduction of regular curricula and special examinations, the reason behind it is, of course, not a suddenly awakened 'thirst for education' but the desire for restricting the supply for these positions and their monopolization by the owners of educational certificates. Today, the 'examination' is the universal means of this monopolization, and therefore, examinations irresistibly advance. As the education prerequisite to the acquisition of

[38] The task consists of taking a triangle of given size and using it to construct a triangle of double size. With a little help by Socrates, the slave finally accomplishes the task (*Meno*, ch. 16, 17, 19).

[39] In contrast, the probably most prominent criticism of this line of reasoning in contemporary practical philosophy is held by Rawls (1971) who understands his second principle of justice – social and economic inequalities should be of the greatest benefit to the least-advantaged members of society; and offices and social positions should be open according to conditions of fair equality of opportunity – as an explicit restriction of offices and social positions *merely* distributed based on individuals' merit (also see Rawls, 1974; Daniels, 1978).

> the educational certificate requires considerable expense and a period of waiting for, full remuneration, this striving, means a setback for talent (charisma) in favor of property. For the 'intellectual' costs of educational certificates are always low, and with the increasing volume of such certificates, their intellectual costs do not increase, but rather decrease.

A selection mechanism that is based on individuals' educational certificates is denoted as a *meritocratic* selection. While Weber's analysis of modern credentialism did not obscure its monopolistic implication, until quite recently, the notion of meritocracy usually had a positive connotation – serving as the benchmark according to which an educational system were to be judged. However, as the next section may show, the demand for *meritocracy* is itself a highly ambiguous concept.

Meritocracy In 1958, the sociologist Michael Young published his satirical novel *The Rise of the Meritocracy, 1870-2033. An Essay on Education and Inequality* (Young, 1958). This "manuscript" pretending to be written in 2033, reports the final victory of the principle of achievement over the principle of ascription. While before individuals obtained their positions in society by assignment or inheritance, now positions are distributed according to individuals' I.Q. and effort. However, while before, talent was distributed almost equally among different groups of society, in the meritocracy, on the one hand, an 'elite caste' of the talented is created, and on the other hand, the untalented form an underclass of known inferiors (cf. Bell, 1972, p. 29f.).

Although Young's *meritocracy* is reported to have broken down in 2034, Bell (1972, p. 30f.) rightly observes that the post-industrial society is, in its self-conception, a meritocracy: Educational certificates as "human capital" (Becker, 1962) serve as passports into the most prestigious positions, while ascriptive factors such as heritage are only an imperfect proxy for an applicant's talent. In this respect, individuals' merit is conceptualized by their tested competence and ability – which is usually operationalized as individuals' IQ or their achievement test results (Hoffer, 2001). The crucial assumption of meritocratic selection is that there is "a close relation between achievement and intelligence and between intelligence and its measurement on the Intelligence Quotient scale" (Bell, 1972, p. 31). To be precise, talent (I.Q.) and merit (achievement) are required to be examined together since "[t]he lazy genius is not one" (Goldthorpe, 1996b, p. 258), while both concepts are supposed to be measurable by standardized tests. The I.Q., in turn, is assumed to follow a bell-shaped distribution, and the top achievers in a particular age category are suspected to be actually the most talented (ibid.).

For about one and a half century, psychometricians have investigated how intelligence could best be measured. Sir Francis Galton (1869) is said to be the first scholar who postulated a construct of general mental ability; and thanks to progress in the development of factor analysis, Spearman (1904) was able to extract a general factor called psychometric g. While Spearman followed a more narrow conception of general ability, Galton understood the general factor more broadly in essentially biological and evolutionary terms (Jensen, 1986). Jensen himself can be denoted to be a follower of the latter tradition, believing in a distinct 'reality' of g apart from its psychometric relevance (Jensen, 1987), and also believing in both the heredity of g and a priori differences (i.e. that cannot be explained by environmental factors such as socialization) in intelligence between ethnicities (Jensen, 1969, 1974; Jensen and Reynolds, 1982; Rushton and Jensen, 2005).

Although the question of test fairness has been an issue discussed for decades (cf. Thorndike, 1971; Linn, 1973), Jensen's position culminated in a monograph titled *The Bell Curve: Intelligence and Class Structure in America* (Herrnstein and Murray, 1994) that addressed the heredity thesis to a broader audience. The book gained a lot of publicity in the US (and still casts its shadow onto public debates in Germany[40]) – but it was also exposed to systematic critique by educational sociologists. While Herrnstein and Murray (1994) aimed to show that intelligence is a far more important predictor of occupational success as environmental factors such as parental socioeconomic status (SES), several authors objected against this conclusion due to several reasons. First, Daniels et al. (1997) provide critique at the thesis that intelligence is predominantly determined by genes. As the authors elaborately discuss, the genes that *potentially* affect IQ are inherited, while IQ itself is not (p. 47).[41] Likewise, the argument implicitly defended by Herrnstein and Murray (1994) that ethnic differences in IQ test scores were due to genetic differences and not to environmental factors is also revealed as being untenable (ibid., p. 62f.; also see Loehlin et al., 1973; Tizard, 1974; Scarr et al., 1977).[42]

Second, Heckman (1995) notes that the g factor was completely overestimated in affecting respondents' social outcomes, and that the view of Herrnstein and Murray (1994) that intelligence could not be manipulated by educational interventions is erroneous since intelligence increases with additional years of schooling (Neal and Johnson, 1996). Regarding the relative importance of intelligence and social backgrounds, respectively, on later social outcomes, both Fischer et al. (1996) and Korenman and Winship (2000) provide extensive re-analyses of the data used by Herrnstein and Murray (1994) and show that if a more realistic (i.e. richer) set of social background variables is considered, intelligence is far from being a more important predictor of later social outcomes than social backgrounds.

Blowing in the same horn as Herrnstein and Murray (1994), Saunders (1997) postulates that individuals' effort and ability outweigh (social) environmental factors in predicting National Child Development Study (NCDS) participants' occupational class at age 33. Furthermore, the fact that in Saunders' (1997) analyses the correlation between ability and class of destination is higher than the one of ability and the class of origin is interpreted as "that the occupational class system is to some extent selecting

[40] In 2010, the former state finance minister of Berlin, Thilo Sarrazin, published a controversial book wherein he sketches a pessimistic picture of the potential consequences of demographic change, a growing underclass and increasing migration to Germany from Muslim countries (Sarrazin, 2010). His discussion about the heredity of intelligence (Sarrazin, 2010, ch. 3) mainly stems from *The Bell Curve* (cf. Sarrazin, 2010, p. 419).

[41] Note that the meta-analysis by Daniels et al. (1997) estimates IQ heritability effects of .34 (in a stricter test) to .48 by maximum (in a weaker test) – which is a far cry from the values of .5 to .8 as declared by Herrnstein and Murray (1994, and naïvely parroted by Sarrazin, 2010).

[42] Goldberger and Manski (1995, p. 771) pointedly summarize the discussion in *The Bell Curve*: "To us, HM's [i.e. Herrnstein & Murray's; *DB*] treatment of genetics and race is akin to standing up in a crowded theater and shouting, 'Let's consider the possibility that there is a FIRE!"'

by ability irrespective of social class origins" (Saunders, 1997, p. 281).

In contrast, Breen and Goldthorpe (1999) provide a re-analysis of the NCDS data used by Saunders (1997) and get to the conclusion that "while merit certainly counts in mobility processes, children of disadvantaged class origins have to display *far more merit* than do children of more advantaged origins in order to attain similar class positions" (Breen and Goldthorpe, 1999, p. 21). Controlling for several covariates, to get the same score on the Hope-Goldthorpe scale of occupational desirability, working-class children would need to outperform middle-class children by approximately one standard deviation in their score in the General Ability Test. Hence, just as in case of the *Bell Curve* debate, this little discussion also ended in a rejection of the attempt of overemphasizing the impact of 'merit' in educational mobility.[43]

In sum, and in the words of Goldthorpe (1996b), the "sociological fantasy" (p. 255) of meritocracy, as introduced by Young (1958), "has engendered both fears and hopes" (p. 280), but "its satirical and critical quality was in fact largely overlooked, or at all events, discounted" (ibid.). From a more critical point of view, one can also call it "philosophically problematic and morally questionable" (Goldthorpe and Jackson, 2005, p. 17).

Solga (2005) characterizes meritocracy as a normative self-definition of modern societies in order to legitimize social inequalities – which is accomplished via five different means (also see Becker and Hadjar, 2011, p. 52-54): i) the 'natural' foundation of social inequalities on biological differences in intelligence or talent (e.g. Jensen, 1969; Herrnstein and Murray, 1994); ii) the depiction of inequality as a functional prerequisite of society (e.g. Davis and Moore, 1945); iii) the necessity of organized educational processes with certificates as signals of individuals' talent; iv) an understanding of social inequality as a problem of individual optimization; and v) an emphasis on the objective character of achievement.

While the first point has been extensively discussed above, and items ii) and iii) would more lead into the field of political philosophy, the next section on the social mechanisms behind the concept of *inequality in educational opportunities* (IEO) will provide a critical assessment of the last two points.

Inequality in educational opportunities On the one hand, without considering environmental factors, the heritability effect on IQ would be overestimated; and on the other hand, even heritability itself is environment-dependent (Daniels et al., 1997). Hence, in this section I will discuss the most influential theories about effects of individuals' social environment on their educational opportunities. In the terminology introduced by Merton (1987) and Goldthorpe (2001), the generative processes of the phenomenon of IEO have to be outlined.

The *Coleman Report* (Coleman, 1966) is one of the first and still one of the most important large-scale studies of social inequality in educational opportunities. While

[43]For additional empirical analyses of the *education-based meritocracy* hypothesis see Goldthorpe (2003) and Goldthorpe and Jackson (2008).

its crucial hypothesis that student socioeconomic status has a higher impact on student achievement than school differences – resulting from differences in resource allocation – has fostered a viral debate (Cain and Watts, 1968; Bowles and Levin, 1968; Coleman, 1968; Moynihan, 1968; Aigner, 1970; Cain and Watts, 1970; Coleman, 1970; Carver, 1975; Eysenck, 1975), the fact *that* students' social backgrounds are an important factor in determining achievement differences also build a main argument in the Wisconsin model of status attainment (Sewell et al., 1969, 1970). Its crucial outcome is respondents' occupational attainment in adulthood, and its explananda are i) respondents' prior educational attainment, ii) their prior aspirations regarding the prospective occupation in the future, iii) their educational aspirations (measured by the intention to attend college), iv) the influence of *significant others* such as parents, teachers and friends, v) the quality of academic performance measured by students' rank in high school class, vi) parental SES, and vii) respondents' mental ability measured by results of a Henmon-Nelson test (Sewell et al., 1969, p. 85). In their path model – which was later exposed to considerable critique since it was accused to reduce causality to correlations (Freedman, 1987; Hedström and Swedberg, 1996) – based on a sample of farm-reared men from Wisconsin first studied in 1957 and re-sampled in 1964, the standardized beta coefficients reveal that the strongest path is leading from mental ability to educational performance (.62), but the next two highest coefficients are significant others' influence on both respondents' educational and occupational aspirations (.45 and .42, respectively). Next in size comes a direct effect of significant others' impact on academic performance (.39), a path of prior educational attainment on later occupational attainment, and a direct effect of educational aspirations on educational attainment (cf. Sewell et al., 1969, table 3).

Shortly later, Sewell et al. (1970) were able to replicate the main findings of their 1969 model based on a more general sample of the Wisconsin data since they had just to add a few more arrows (e.g. from mental ability to significant others' influence) to adapt the model to the extended sample. Following this revised version of the Wisconsin model, students' aspirations and expectations became the central mediating variable in status attainment research (Morgan, 2006, p. 1529). While the spread of Bourdieu's theory of cultural capital (Bourdieu, 1973) in a way counterbalanced the stream of research presuming a direct association between students' aspirations and their educational attainment (Morgan, 2006), another theory still casts its shadow onto contemporary research on inequality in educational opportunities: Boudon's distinction between *primary* and *secondary* effects of social inequality.

In 1974, Boudon published his monograph about *Education, Opportunity, and Social Inequality: Changing Prospects in a Western Society* (Boudon, 1974). While some of its conclusions regarding the relationship between *inequality of educational opportunities* (IEO) and *inequality of social opportunities* (ISO) were exposed to various critical remarks (Hauser, 1976; also see Boudon's (1976) reply), Boudon's distinction between *primary* and *secondary* effects of social inequality became one of the most influential concepts in contemporary quantitatively-oriented IEO research (see e.g. Meulemann, 1979; Breen and Rottman, 1995; Erikson and Jonsson, 1996a,b; Goldthorpe, 1996a; Breen and Goldthorpe, 1997; Müller-Benedict, 1999; Becker, 2000; Breen and Jonsson, 2000; Need and De Jong, 2001; Solga, 2002; Becker, 2003; Erikson et al., 2005; Becker and Schubert, 2006; Jaeger, 2007; Müller-Benedict, 2007; Pietsch and Stubbe, 2007; Stocké, 2007; Van De Werfhorst and Van Tubergen, 2007; Maaz et al., 2008; Schneider, 2008; Becker and Hecken, 2009a,b; Breen et al., 2009; Kloosterman et al., 2009; Erikson and Rudolphi, 2010; Glaesser and Cooper, 2011; Hillmert and Jacob, 2010; Neugebauer, 2010; Schubert and Becker, 2010; Tolsma et al., 2010; Karlson and Holm, 2011; Morgan, 2011; Modin et al., 2013; Morgan and Winship, 2012; Hillmert, 2013).[44]

The *primary* effect of educational inequality states that the lower educational success of lower-SES children may be due to their lower capabilities – be they defined as educational interests, intellectual skills, effort or motivation (Müller-Benedict, 2007). *Part* of the primary effect may indeed be genetic in the sense of Jensen (1969) and Herrnstein and Murray (1994), but another, presumably greater part of the above-mentioned characteristics is acquired during socialization (Erikson and Jonsson, 1996a, p. 10f.).[45]

The *secondary* effect, contrarily, operates via stratum-specific differences in educational decision making due to differential opportunity-cost structures, and Boudon's crucial assumption is that secondary effects still take place once primary effects have been controlled for (Nash, 2005). In Boudon's (1981, p. 191) words:

> "The subject's class of origin (or the class to which a family now belongs) will crucially affect his choices of one or the other option. If their current success is mediocre, the family unit will consider itself 'satisfied' if the child has reached an academic level enabling him to aspire to a social status equal or higher than his own, even if this status is not especially high. A well-placed family unit will on the other hand strive (I ought to add: more often than not) to 'push' the child so that he doesn't fail (even if he doesn't enjoy a greater success)."

As mentioned, the idea that aspirations may vary by social class has not been invented by Boudon. Actually, already Hyman (1953) postulated that lower-class individuals aspire lower aims than higher-class individuals (which he attributes to class-specific value systems), and Keller and Zavalloni (1964) respecified his approach by introducing the idea of class-specific relative distances towards particular values. However, while Keller and Zavalloni (1964, p. 60) on the one hand understand social class as an "intervening variable between individual ambition and social achievement", and on the other hand do not entirely discard the value-relatedness of aspirations, Boudon (1974) overcomes this shortcoming by first modeling aspirations as a social mechanism that is located *between* social class and (educational) achievement; and second, by relating class-specific differences in aspirations to differences in utility considerations (Erikson and Jonsson, 1996a, p. 13f.).[46] Hence, Boudon not only argues against both reductionist covering-law

[44]For further reviews of the literature see Kristen (1999); Stocké (2010); Becker (2012); Solga and Becker (2012).

[45]In contrast, see Lucchini et al. (2013) for a recent contribution that once again strives to take up the cudgels for a genetic explanation of IEO. Note, however, that their variance decomposition approach provides nothing more than a black-box explanation (which the authors themselves partly acknowledge; cf. Lucchini et al., 2013, p. 5) without revealing the particular mechanisms behind.

[46]Note, however, that Erikson and Jonsson (1996a, p. 28f.) still observe a value assumption in Boudon's theory, "namely that, given the social distance traveled, the negative effect of social demotion on benefits is higher than the corresponding positive effect on social ascent".

type variable sociology (Boudon, 1976, 1979) and a too narrow notion of rationality (Boudon, 1996, 1998, 2003), his own theory on social inequality in educational opportunities (Boudon, 1974) also is a prominent example of a *mechanism-based explanation* opening the black box of social background effects in order to get to a better *understanding* of IEO in the Weberian sense.[47]

The idea that utility considerations may shape students' (or their parents') educational decisions was further elaborated by Goldthorpe (1996a).[48] Referring to Boudon's elaborated rational action theory – and, implicitly, also to the *maxim of Occam's Razor* (Thorburn, 1918) – Goldthorpe (1996a, p. 490) argues that it is simpler (i.e. more parsimonious) to assume that there is no class-specific variation in either aspirations or in potentially underlying value systems. Instead, Goldthorpe develops the idea of regarding education as an investment good the costs and benefits of which vary by social classes.

Goldthorpe's simple premise is that each family will strive to avoid downward mobility. Unsurprisingly, for lower-educated parents, this goal will be reached already for lower educational qualifications of their children – while for higher-educated parents, a far higher degree will have to be obtained. Moreover, for the offspring of parents in less advantageous positions, each failed attempt of trying a higher educational alternative will be more serious in its consequences concerning both monetary (earnings foregone; loss of financial support) and transactional costs (a loss in itself; the risk of dropping out of the educational system).

The utility model of students' educational transitions was first formalized by Erikson and Jonsson (1996a). They introduce a simple 3-parameter model (which heavily resembles Goldthorpe's theoretical considerations) postulating that students' utility is affected by educational benefit B, costs of education C, and the expected probability of educational success P. B can also include prospective benefits during academic studies, and C comprises of both monetary and psychological costs of education. When in case of educational success, educational utility consists of educational benefit net of costs, and in case of failure, only educational costs remain, the utility model reads (Erikson and Jonsson, 1996a, p. 14):

$$U = (B - C)P - C(1 - P) \Leftrightarrow U = PB - C. \tag{1}$$

Thus, a student's utility equals her expected educational benefit times the expected probability of educational success minus expected costs (ibid.). As the authors empha-

[47]One could argue that in contrast to Boudon (1974), Herrnstein and Murray (1994) fall back into a mechanical push-pull explanation (Machamer et al., 2000, p. 2) in terms of a suspected primacy of heredity.

[48]Much of what followed with and after Goldthorpe (1996a) was already sketched in Meulemann (1979). A difference between Meulemann (1979) and the later models by Goldthorpe (1996a); Erikson and Jonsson (1996a,b); Breen and Goldthorpe (1997), and Esser (1999) that is crucial for the objective of theoretical identification of self-fulfilling prophecies in section 4.3 is first that Meulemann (1979) does not provide a formal theoretical model, and second – and more important –, that he explicitly excludes subjective expected probabilities of success and thus assumes all cost-benefit terms to occur with a probability of one (Meulemann, 1979, p. 399, footnote 5).

size, two educational alternatives i and j can yield the same expected educational utility ($P_iB_i = P_jB_J$) if $P_i < P_j$ and $B_i > B_j$. However, a more *risk averse* person would always opt for alternative i, notwithstanding the higher expected benefit of alternative j.

Breen and Goldthorpe (1997) specify the model by Erikson and Jonsson (1996a) more precisely by introducing three educational outcomes on the one hand and class-specific status destinies on the other hand. While a presentation of the exact mathematics of their model is beyond the scope of this introduction, their main idea should be captured though: Even if continuing in education was not dependent on the costs of remaining at school, and if there were no class-specific ability differences, service-class students would still be more likely to continue a high level of education than working-class students.

However, since evidence on *primary* effects of social inequality reveals class-specific differences in both academic ability and resources, and students' subjective expected probability of successfully continuing the chosen school track can be assumed to be endogenously influenced by their academic ability, differences in transition propensities between the social classes will be further broadened than a more parsimonious explanation based on simple status-maintaining utility consideration would suggest.

Hartmut Esser (1999) basically builds on several elements of both Erikson and Jonsson's (1996a) as well as Breen and Goldthorpe's (1997) respective models, but one the one hand, he introduces additional parameters (such as the impact of the expected status decline, once a chosen school track was not finished with success, on parental decisions), and on the other hand, he splits up the extended model into two smaller components, which allows him to get to the following conclusion: A student opts for the higher educational track if the expected *educational motivation* – consisting of the expected educational benefit and the expected status decline (times its impact on parental decisions) – exceeds the estimated *investment risk* – defined as the expected costs of education weighted by the inverse of the subjective probability of educational success.

Whether one favors a more parsimonious model, such as the initial proposition by Erikson and Jonsson (1996a), or a more complex explanation as the one by Esser (1999): The crucial point to be made here is that all of the above theoretical transition models open the black box by introducing simple utility-based assumptions as the underlying social mechanisms in order to obtain a better *understanding* of social inequality in educational opportunities. In contrast to critique as it had been put forward against the early Wisconsin status attainment model (Hedström and Swedberg, 1996), significant correlations or regression estimates cannot be confounded with causality since the respective indicators will always have to be linked back to the underlying utility assumptions that should account for the empirical results.

Dissecting the assumptions of the rational choice (or *s*ubjective *e*xpected *u*ility = SEU) model of parental educational decisions according to the terminology introduced by Hedström (2005) nicely illustrates its advantage regarding parsimoniousness: While parents' *desires*, i.e. their absolute educational and occupational aspirations are assumed to be constant among classes (Keller and Zavalloni, 1964; also see Meulemann, 1979, 398, footnote 4), it is their *beliefs* about the expected benefit of education, the perceived amount of status decline, or the subjective expected probability of educational success that should be different among the social strata. Hence, the above-cited theoretical accounts all assume a *belief-mediated* social mechanism to underlie educational transition decisions.[49]

Another issue to emphasize at this point is the question of the underlying *decision heuristic*. As outlined above, the rationality principle itself is "almost empty" (Popper, 1994, p. 169), and we have to expect that simply referring to actors' rationality without giving additional specifications will not be the end of the story in the development of a satisfying mechanism-based explanation (also see Gambetta, 1998, p. 118). As it is already implicitly sketched in Boudon (1974), a bit more overtly supposed by Erikson and Jonsson (1996a) and Goldthorpe (1996a), and explicitly discussed by Breen and Goldthorpe (1997) and Esser (1999), a crucial assumption of a utility-based transition model is actors' *relative risk aversion*. However, as Esser (1999, p. 274) highlights, we should not assume that relative risk aversion is a mechanism counterbalancing the 'inner logic' of an SEU explanation of educational transitions. Contrarily, relative risk aversion rather *follows* from straightforward calculus according to the SEU rules as a consequence of the social situation's opportunity structure. Hence, the underlying decision heuristic reads "Maximize the subjective expected value!" equally for all social strata – while, as outlined, corresponding subjective expected probabilities should of course expected to vary by social class.[50]

Still, some controversy remains whether primary *or* secondary effects are more important in explaining IEO. By means of simple frequency table analysis, Nash (2005) questions scholars who emphasize the prevalence of secondary effects, and he himself concludes that primary effects are the comparably stronger ones. Contrarily, Jackson et al. (2007) apply the method of *counterfactual analysis* of primary and secondary effects that has been introduced by Erikson and Jonsson (1996b) and extended by Erikson et al. (2005). The crucial idea of this method is that both kind of effects can be disentangled by relating students' achievement distributions and transition propensities, respectively, to the appropriate linking functions. Concretely, students' class-specific academic performance is modeled by the area under its normalized distribution (i.e. the integration of the standard normal density function with students' class-specific mean performance and respective variance as its identifying parameters); and class-specific transition propensities are approximated by the area under the respective logistic regression curves (Goldthorpe and Jackson, 2005; Jackson et al., 2007). Both areas multiplied – which has to be done via numerical integration – allows to estimate, say, the counterfactual transition rate of the underclass *if they had the performance distribution*

[49] Recent evidence by Stocké (2011) aims to challenge the assumption of constant aspirations among social classes: The author finds that in working class, the difference between *idealistic* (i.e. most wanted) and *realistic* (i.e. most probable) aspirations (Gottfredson, 1981) is higher than in service class (also see Stocké, 2009). However, Stocké's analyses did not follow either one of the above-reviewed utility frameworks, and I do not see how realistic aspirations should differ from students' subjective expected probability of successfully completing a given school track – which would still suite the (more parsimonious) assumption of constant desires and varying beliefs among social strata.

[50] As opposed to the above-described relatively parsimonious decision heuristic, in the conclusion section of this introduction, I will propose to investigate the consequences of relaxing Occam's razor in adding the auxiliary assumption of *class-dependent decision heuristics*.

of the salariat, for example.

Decomposing primary and secondary effects in this manner shows that the latter reinforce the former to a considerable extent in that secondary effects account for at least a quarter up to a half of the variance of students' actual transitions. Hence, narrowing the focus merely on primary effects would be both myopic and perhaps lead to ineffective policy conclusions (Jackson et al., 2007, p. 224).

However, Jackson et al. (2007, p. 224) also note that secondary effects of social inequality may also occur "in conjunction, perhaps, with parents, teachers and peers" – which is what the Wisconsin model would have called students' *significant others* (Sewell et al., 1969, 1970). After a short interim conclusion, the next section will summarize the four papers providing both theoretical (in terms of mechanism-based explanations) and empirical evidence (in terms of results from quantitative analyses) that teachers' evaluations *are* a factor that may affect students' transition propensities apart from the parameters of the utility-based theoretical accounts hitherto applied.

Interim conclusion Starting with Weber's (2000) definition of social class in terms of life chances, the preceding section first discussed the *meritocracy thesis* (Young, 1958; Bell, 1972) which states that individuals' social position should depend on their respective *merits* – defined as IQ + effort – and its influence on subsequent discourses about the heritability of intelligence and its impact on individual achievement compared to social background factors (Jensen, 1969; Herrnstein and Murray, 1994). Since most of the supposed evidence on the primacy of intelligence over social backgrounds was revealed to be both incomplete and misleading (Heckman, 1995; Fischer et al., 1996; Daniels et al., 1997; Korenman and Winship, 2000), the focus of the remainder of this section lay on the *generative processes* (Goldthorpe, 2001) of inequality in students' educational opportunities (IEO). Following a classic though still influential educational transition theory (Boudon, 1974), I distinguished *primary* effects (partly due to genetics, but mainly due to social-environmental factors) from *secondary* effects of social inequality due to different utility considerations regarding educational decisions. In its more recent and more elaborated versions (Erikson and Jonsson, 1996a; Goldthorpe, 1996a; Breen and Goldthorpe, 1997; Esser, 1999), more fine-grained bridge assumptions have been added to the initial transition model that allowed me to reconstruct transition differences by social strata as a *belief-mediated* social mechanism (cf. Hedström, 2005) – while actors' desires (i.e. their 'idealistic' aspirations; cf. footnote 49) as well as their underlying decision heuristics – maximize the subjective expected value – are assumed to be invariant among social classes. Now the ground is set for outlining the generative processes – in terms of social mechanisms – approximating the causes and effects of teachers' evaluations as one of students' *significant others* in their educational transition process.

4 Applications

In this section, I will try to reconstruct the social mechanisms implied in the four papers of this volume. The first paper, *Teachers' Evaluations and the Definition of the Situation* *in the Classroom* (chapter II), provides a path model in the tradition of the classic Wisconsin status attainment theory – only that we do not cover the longitudinal aspect and restrict our analyses to the emergence of 10^{th} class teachers' evaluations.[51] Here, I will elaborate more intensely on the *Model of Frame Selection* that could only be sketched in the paper's theoretical section.

The second paper, *Intelligence and Academic Achievement as Contextual-Level Predictors of Teachers' Evaluations: Big Fish Little Pond or Reflected Glory Effect?* (chapter III) places the outcome of teachers' evaluations in the context of reference-group research and asks whether there is a positive or a negative effect of class-level ability and achievement, respectively, on the fact whether a student obtains a 'good' or a 'bad' evaluation by her teacher. Some more words have to be said about to what extent the mechanisms assumed here differ from what is hypothesized in conventional *big fish little pond* research.

The third paper, *The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course* (chapter IV), asks if 10^{th} class teachers' evaluations may induce self-fulfilling prophecy effects regarding students' later educational transitions. Here, I will come back to the idea of secondary effects of social inequality as a *belief-mediated* social mechanism in order to identify self-fulfilling prophecies by changes in subjective probabilities of educational success.

Finally, the fourth paper, *Does the Effect of Teachers' Expectations on Students' Educational Opportunities Decrease Over Educational Transitions? A Statistical Matching Approach* (chapter V) tests whether the phenomenon of decreasing background effects over educational transitions also holds for teacher expectancy effects. The reconstruction of self-fulfilling prophecies as belief-mediated mechanisms is amended by the idea of *Bayesian learning* to account for changes in individual success estimates over time.

4.1 Paper 1: A Frame Selection Model of Teachers' Evaluations

As mentioned above, the structural form of the paper *Teachers' Evaluations and the Definition of the Situation in the Classroom* (chapter II, co-authored by Klaus Birkelbach[52]) is to some degree comparable to the Wisconsin status attainment tradition – only that one generalized member of the students' *significant others*, i.e. their teachers (and in particular, teachers' evaluations) is the focus of the analyses. It was already outlined in section 2.2 that a rather criticizable point of the Wisconsin model (and of path models in its tradition) is that it stands at risk of mixing up (significant) correlations with causality. Although path model luminaries such as Jöreskog (1993) heavily emphasize that one should only allow for additional path coefficients or covariances that also make sense from a theoretical point of view, the researcher is often tempted to relax constraints and sacrifice methodological rigor for empirical model fit gains (so-called

[51] Of course, by our theoretical foundation, we also have the demand of not to fall into the trap of theoretical underspecification that has been objected against the Wisconsin path models (Freedman, 1987; Hedström and Swedberg, 1996).

[52] A great deal of the theoretical considerations in this subsection is also joint work with Klaus Birkelbach.

post-hoc model fitting; also see Byrne et al., 1989).[53] Hence, it will be necessary to outline some more *fine-grained* explanations of the suspected causal structure than it was possible in the paper.

The dependent variable of interest are 10^{th} class teachers' evaluations whom of their students they consider to be able for academic studies – and whom of them not. Teachers were asked to evaluate their students in the teacher survey of the *Cologne High School Panel* (1969), and we make use of additional student and parent questionnaires to explain how teachers might arrive at a positive or negative judgment of their students.[54]

In the paper, we argue that in order to arrive at a more systematic action-theoretical model of teachers' evaluations, it is useful to start with the general model of sociological explanations (Coleman, 1990; Esser, 1993a) that links the conditions and alternatives of a social situation on the macro level (*the logic of situation*) to actors' expectations and evaluations on the micro level. The latter, in turn, shape individuals' action (which is referred to as the *logic of selection).* In a third step, individual actions are aggregated to a new social situation via transformation rules (the *logic of aggregation*).[55] The crucial assumption of the logic of selection is that this step operates via a mechanism of *frame selection* (Esser, 1993a, 1996b) which assumes that an actor sequentially defines a social situation by selecting a particular *frame*, while she arrives at a concrete action by recurring to a more specific *script*. Whereas Boudon (2003) is critical of the usefulness of explanations by frame selection theories – regarding them even as equal to "introducing a black box" (p. 7) –, we here follow Opp (2011, p. 218) who tends to agree that the MFS might be applicable to explain actors' preferences as well as their beliefs – whereby it would suit the general framework of mechanism-based explanations (Hedström, 2005). A second argument for this step comes from the *pragmatist theory of mechanisms* by Gross (2009) who virtually aims to oppose against rational choice theory in general which is too narrow from his point of view. Instead, Gross emphasizes that human action "(...) involves an alternation between habit and creativity. (...) Only when preexisting habits fail to solve a problem at hand does an action-situation rise to the forefront of consciousness as problematic" (Gross, 2009, p. 366). Grounding on these considerations, a pragmatist theory of social mechanisms is given "*as composed of chains or aggregations of actors confronting problem situations and mobilizing more or less habitual responses*" (Gross, 2009, p. 368; orig. emph.). While apart from the use of analytically imprecise phrasings such as 'chains or aggregations', I would in general agree that the processes of action they describe are suitable to be subsumed beneath the concept of social mechanisms, the exemplifications below might show that the idea

[53] A caveat against path analysis is also held by Freedman (1987) who objects that i) path analysis usually assumes a linear relationship between the variables of interest – which may be violated –; ii) the causal ordering of the variables has to be thoroughly deduced from theory *before* model fitting; and iii) even in the most complex models, omitted variable bias may still distort the estimates since specification of an exhaustive empirical model will hardly be possible.

[54] For more detailed information on the data see the respective sections in the four papers.

[55] See Hedström (2005, ch. 6) as well as Bornmann (2010) for a more elaborate discussion of the possibilities of *agent-based modeling* in order to get an intuition about the consequences of different mechanisms of aggregation.

of an 'alternation between habit and creativity' should not be viewed as standing in opposition to rational choice theory in general. Instead, it suits the assumption of *variable rationality* as hold by the *Model of Frame Selection* (see below).

In accordance with the literature, we deduce three distinct dimensions that could affect teachers in shaping their evaluations regarding 10^{th} class students' prospective aptitude for academic studies: students' academic performance, their cognitive ability, and their social backgrounds (which are later split up into socioeconomic status and aspirations). The question is now how these theoretical concepts can be linked to teachers' evaluations under the assumption that the latter emerge by means of a more or less rational selection process.

In order to answer that question, it is fruitful to refer to Esser's and Kroneberg's enhancement of Kahneman and Tversky's (1984) early version of the framing approach towards a general theory of action (Esser, 1996a, 2010; Esser and Kroneberg, 2010; Kroneberg, 2005, 2006; Kroneberg et al., 2008, 2010; Kroneberg, 2011; Kroneberg and Kalter, 2012). The *M*odel of *F*rame *S*election (MFS) first separates individual action analytically into *frame selection*, *script selection*, and *action selection*. Second, it assumes that on each level, the selection process may vary between an unconscious automatic-spontaneous processing, and a 'rational' penetration of the respective selection stage. By doing so, the MFS synthesizes arguments from both the interpretative paradigm and rational choice theory to account for what is referred to as actors' *variable rationality* (Kroneberg, 2006).

The *frame selection* refers to actors' individual definition of the situation. Before action may take place, an actor has to answer the question, "What is going on here?" (Goffman, 1974, p. 8). By explicitly considering actors' *individual* definition of the situation, the MFS also builds on assumptions of symbolic interactionism and actors' *Lebenswelt* (also see Esser, 1993b) and thus holds the demand to arrive at an *understanding* explanation (Kroneberg, 2011, p. 120). According to that theory, in most cases, the actual situation is defined in an automatic-spontaneous mode (*as-mode*), depending on the match of the actor's perceptions with internally stored mental models. The match is determined by (1) the frame's general *availability* (determined by socialization and life course experiences) (2) the perceivability of unique situational *objects* (such as cultural symbols and gestures), and (3) the mental *link* between situational objects and a specific frame (e.g., *Do 20 seconds of silence count for 'communication breakdown'?* – which may be subject to cross-cultural variance; see Kroneberg, 2011, p. 131). Only in cases without such a match, a reflecting-calculating definition of the situation (*rc-mode*) is required.

Once a particular situation has been defined, more concrete *scripts* of action reduce the complexity of possible alternatives of actions. Same as frame-selection, script-selection varies between an automatic activation of available scripts – acquired through the process of socialization and depending on both the internalization of norms and the habitualization of routines (*as-mode*) – and a rational reflection about the alternatives at hand (*rc-mode*) as well.

And finally, individuals' concrete *actions* may vary between *as-mode* and *rc-mode* processing as well – depending on whether a particular script exactly demands a subsequent action, or if the consequences of different alternatives are explicitly considered. In the latter case, SEU theory would be an adequate specification.[56]

On each level, the match between the social situation and its mental frame determines which form of processing is intuitively chosen. If the actor's definition of a social situation is without any doubts, and a chosen frame is strongly linked to a particular script, and this script requires certain action(s), then the *as-mode* is the adequate since most efficient coping strategy. However, if there *is* 'definitional complexity', a more rational penetration of the social situation as well as a more conscious selection of scripts and individual actions might be more conducive. Exactly this is meant with the notion of actors' *variable rationality* (Kroneberg, 2005, 2006).[57]

As regards teachers' evaluations as they had been surveyed in the *C*ologne *Hi*gh *S*chool *P*anel (*CHiSP*), the frame of the underlying social situation should be rather unambiguous: the demand of an anonymous, non-binding assessment of students' future academic potential. Thus, teachers should recognize the demands of this situation more or less automatically (*as-mode*).

Now the question arises which *scripts* are at the teachers' disposal in this situation of a non-binding assessment. The answer given in the paper is that this particular frame requires a script of professional pedagogic diagnostics. As already sketched above, we assume that as long as teachers' evaluations are grounded on meritocratic criteria like students' academic performance; they follow occupational standards that are deeply-rooted in every teacher's mind. Thus, evaluations which are based on meritocratic criteria will emerge rather automatically in line with the *as-mode*. However, though probably most legitimate, these criteria will be not the only ones determining teachers' actual evaluations. In total, it is possible to enumerate three different types of processing that might come into play besides meritocracy.[58]

First, Bourdieu (1986) developed a theory on different forms of capital in which upper-class students' *habitus* – defined as a system of dispositions (socially acquired schemes of perception, thought and action that are stable over time) – almost perfectly matches with the habitus of their teachers who usually originate from the same social stratum and thus have incorporated a similar system of social dispositions. This positive social discrimination of upper-class students is twofold: On the one hand, upper-class students are usually more familiar with codes (or routines) that are necessary to acquire the cultural goods that are taught in class. On the other hand, these first-order codes depend, in turn, on second-order codes of perception, communication and self-control

[56] Chess is a good example why it is useful to introduce an analytical separation between script selection and selection of action: The frame of a chess game should be unambiguously clear (*as-mode*), so is the required script of chess rules (*as-mode*). However, this script still allows a multitude of permitted maneuvers – which is why the selection of action will take place in *rc-mode*.

[57] The *mode-selection* that decides about the mode of each analytical level is assumed to be a preconscious process but uses SEU theory *as a heuristic* (Esser, 2010; Kroneberg, 2011, p. 144f.).

[58] In section 3, I critically discussed the "fears and hopes" (Goldthorpe, 1996b, p. 255) that have been associated with the idea of meritocracy. Consequently, the idea of the *meritocracy-as-mode* is not that the authors of the underlying paper personally favor this mode in normative terms, but that the idea of a meritocratic evaluation might be regarded as most legitimate by educational decision-makers (including teachers).

strategies acquired in socialization that may affect even factors like students' motivation and aspirations (Bourdieu, 1986; Bourdieu and Passeron, 1990; also see Kroneberg, 2011, p. 105f.). Thus, upper-class students with higher cultural capital will not only dispose of more knowledge of school-relevant contents, but they will also be more able to perceive and to communicate according to norms and via symbols that come up to the expectations of their teachers (also see Dumais, 2006, p. 85f.).[59]

We do not follow all implications of the *habitus* concept in the sense of Bourdieu's original idea. On the one hand, as Lareau and Weininger (2003) have pointed out, much of the usage of cultural capital as a concept in educational research can be traced back to an early DiMaggio (1982) paper. As a result of the reception of this paper, a narrowing of cultural capital operationalization by 'highbrow' culture as well as a distinct treatment of cultural capital and 'achievement' can be noted – of which both was beyond Bourdieu's intention. Second, and more important, Goldthorpe (2007) highlights that Bourdieu's *habitus* concept cannot be separated from his conflict-theoretical approach towards schooling in general. According to this view, cultural capital per se is arbitrary and only 'used' by teachers in school in order to maintain the current social structure by means of a social closing mechanism. Furthermore, a *habitus* once acquired in family has to be regarded as stable and cannot be changed by means of schooling interventions (which would, according to Bourdieu's conflict-theoretical approach, not even be aspired by teachers at all).

Therefore, in the following, the idea of a *habitus* effect should be understood rather metaphorically to refer to comparably unconscious, but *habitual* social status effects and without the ideological burden initially intended by Bourdieu. While at a first glance, one might ask which conceptual gain would be arrived at compared to Weber's dictum of class- and status-related conduct of life (*ständische Lebensführung*) in his definition of social class (Weber, 1978, p. 306f.), our argument is that the idea of *variable rationality* (Kroneberg, 2005, 2006) is a more general concept that also subsumes status-related conduct of life. Also to show this, both acquisition and efficacy of potentially unconscious status-related 'dispositions' should be analyzed within the framework of the MFS.[60] As Kroneberg (2011, p. 104-108) remarks, Bourdieu's thesis that action appearing to be rational from an observant's point of view can be rooted in unconscious, long-time dispositions suits the assumption of *variable rationality* when a 'disposition' in Bourdieu's sense is detached from action-theoretical determinism. Instead, Kroneberg (2011, p. 107) argues in line with several other authors (Elster, 1983; Hedström, 2005; Yaish and Katz-Gerro, 2012) who offer a reading of Bourdieu's *habitus* that can be referred to as 'brushed against the grain' (Benjamin, 1968, p. 257) in allowing also more

[59] Also Erikson and Jonsson (1996a, p. 24) highlight that evidence on the impact of (parental) cultural resources on (students') differences in educational transition probabilities between the social strata is often indirect – which legitimizes the use of *habitus*-alike processes as an unobserved mechanism. That is, even if indicators for *primary* and *secondary* effects of social inequality are controlled for, parental socioeconomic status might still reflect habitus-related attributes also affecting teachers' evaluations.

[60] For an MFS-related account to the above-mentioned 'highbrow' cultural practices of individual actors see Weingartner (2013).

conscious reflexions *given* certain status-related action constraints. Kroneberg quotes Bourdieu's assertion that *habitus* offers a "conditioned and conditional freedom" that is "as remote from a creation of unpredictable novelty as it is from a simple mechanical reproduction of the initial conditionings" (Bourdieu, 1977, p. 95).

Transferring this on the emergence of teachers' evaluations, *as long as* the above-mentioned match of symbolic codes *unconsciously* influences teachers in their evaluations (as in common-sense *habitus* reception), this, too, is in line with the *as-mode* of automatic processing. As Kroneberg (2006, p. 18) points out (also see Kroneberg, 2011, p. 132f.), there will be greater activation of an *as-mode* script

- the higher its general *availability*,
- the higher its *accessibility* given the selection of frame, and
- the higher the *match* of the selected frame.

The *availability* of a script describes how strongly it is mentally anchored, and its *accessibility* represents the degree of mental association between frames and scripts.

In our case, the *as-mode* prevalence of students' social background criteria in a sense conventionally referred to as *habitus* effects will particularly depend on the script's availability, i.e. "how strongly an actor has internalized certain norms or become[s] accustomed to certain routines" (Kroneberg, 2006, p. 18). The main point here is that in accordance with the mode of automatic processing, actors do not have the opportunity to select between different *as-mode* scripts deliberately; instead, there is always only one dominant *as-mode* script – whether it approximates more to the ideal type of meritocracy or more to teachers' habitual recurrence of students' social backgrounds. In sum, the first possible deviance from the *as-mode* meritocracy model would be a more or less pronounced (but still unconscious) shift towards teachers' habitual consideration of students' social backgrounds which presumably anchor the shaping process of teachers' evaluations as well. To keep a well-established label, but without buying Bourdieu's conflict theory, we refer to this particular script as *habitus-as-mode*.[61] *Second*, however, and in line with the assumption of actors' *variable rationality*, the extent to which teachers' recurrence on student social background criteria merely follows a *habitual*, i.e. unconscious automatic selection process (*as-mode*) may vary as well. Teachers' might consciously take into account that *apart from* their current academic achievement, some students might be more able to both start and successfully complete academic studies due to their more favourable social backgrounds: On the one hand, teachers might suppose that students from the higher social strata dispose of characteristics making them more 'suitable' to higher education. This *could* be habitus-alike socially acquired

[61] Another theoretical account that conjoins the efficacy of symbolic codes with models of situational framing is the one proposed by Bernstein (1971, 1981) – who himself notes a certain theoretical proximity to Bourdieu's concept of habitus (Bernstein, 1990, p. 3). For a more elaborate discussion on the similarities and differences between Bernstein and Bourdieu see Bourdieu (1991, p. 53), Harker and May (1993) as well as Bernstein's (1995) reply to Harker and May. For an empirical test of Bernstein's theory see Meulemann (1976).

schemes of perception, thought and action that match those of their university teachers – but in contrast to the *habitus-as-mode*, school teachers more or less *consciously* reflect on this potential later match/mismatch. On the other hand, school teachers might simply expect that upper-class parents, let's say having an academic background themselves, would be more able to support them. Thus, in that case, the dominant script that follows the *as-mode* framing of the social situation would be a mixture of an *as-mode* assessment of students' academic performance and of a *rc-mode* evaluation of the estimated impact of students' social backgrounds on their potential academic success at university.

Having allowed for variable rationality with regard to the *efficacy* of social backgrounds, the question of the *acquisition* of status-related dispositions – whether taking effect consciously or unconsciously – remains unanswered. Above, I already followed Goldthorpe (2007) in criticizing Bourdieu's rather mechanistic understanding of socialization. Whereas even the most detailed elaboration of the MFS (Kroneberg, 2011) remains silent on this issue[62], my proposition is to allow for variable rationality also concerning the learning of habits. While early behaviorist learning theories mainly assumed a simple stimulus-response model of social learning, Bandura (1969) introduced an idea of social learning that could also be described in terms of the mechanism of *rational imitation* (Hedström et al., 1998; for an overview on socialization theories see Maccoby, 1992).

Excursus: Variable rationality of habit learning

As Bandura (1969, p. 220) points out, conditioning theories cannot explain "how new response patterns are acquired observationally, particularly under conditions where an observer does not overtly perform the model's responses during the acquisition phase", i.e. the behavior learned by observing is performed much later. Therefore, he postulates that learning involves both an imaginal and a verbal representational system that are mentally stored and mediate future behavior. As these kind of *retention processes* are, amongst others, reinforced by the role model's social status, the applicability of Bandura's account on socially acquired schemes of perception, thought and action such as Bourdieu's *habitus* is straightforward. Even more, Bandura (1969) opens the black box of simple conditioning in specifying several mediating mechanisms of this long-term acquisition process in terms of punishment/reward, peer-group influence, social mobility, institutions (e.g., the school system) or even media consumption.

Furthermore, Bandura (1969) objects against simple *mimicry* learning models in favor of those open for development of novel patterns of behavior. While doing so, children may keep those forms of behavior that have proven useful in order to achieve certain goals, and discard other forms that have not (also see Becker, 2012, p. 56f.). My argument is that Bandura's specification of intervening mechanisms and his demand to account for cost-benefit-related maintenance of novel forms of behavior transcends simple *as-mode* socialization theories (as the one by Bourdieu, 1986) and approximates towards *rc-mode* forms of rational imitation.

A recent model combining both *as-* and *rc-mode* learning was proposed by Sun et al. (2005): The authors separate the cognitive structure into a 'bottom level' consisting of *implicit* knowl-

[62] Although some passages in Esser (1999, ch. 9) touch on the idea of learning by both conditioning and rational reflection, this was before the idea of variable rationality was further elaborated on – which is why I still see the need for a more thorough specification.

edge responsible for automatic memory-based processing, and a 'top level' consisting of *explicit* knowledge responsible for explicit hypothesis testing. Sun et al. (2005) develop a complex formal model on the conditions of implicit and explicit learning that cannot be reviewed here, but its main idea is straightforward: Analogous to human action, also learning in general (and socialization in particular) may take effect via both unconscious, implicit conditioning (*as-mode*), and via conscious, explicit rational imitation or hypothesis testing (*rc-mode*).

Hence, providing an MFS-related explanation of how students' social backgrounds might affect teachers in shaping evaluations on their students requires to introduce variable rationality also in socialization theories. Status-related perceptual schemes may be acquired both automatic-spontaneously (*as-mode*) and by means of rational imitation (*rc-mode*). In either case, these dispositions may influence both unconsciously (*habitus-as-mode*) and by means of rational reflection (*habitus-rc-mode*) later individual action.

Apart from social backgrounds, supplemental to an *as-mode* assessment of students' academic performance, *third*, teachers might refer to additional criteria of students' general academic ability like their (estimated) intelligence or motivation. Apart from the most visible academic performance of the students (usually operationalized by their school grades), teachers could find rational reasons for differences in ability that might affect students' success probabilities but are not reflected in grades. Students with the same grade might differ in cognitive abilities or in the motivation they invested to achieve this grade, and these differences might also lead to differences in their (estimated) probabilities of university success. Our main point here is that in contrast to the *as-mode* assessments of students' academic performance, we assume teachers' additional considerations about students' ability to be the result of rational reasoning (*rc-mode*).

These theoretical considerations can be summed up as follows: Teachers' evaluations as measured in the *CHiSP* data emerge in a social situation that is framed more or less automatically *(as-mode*) by the teachers. In a second step, teachers' actual decisions, i.e. in DBO terminology their *beliefs* about certain students, will be formed according to a specific script of action which may vary between an automatic (*as-mode*) and a rational (*rc-mode*) pole of information processing. In the most probable script of action, teachers intuitively ground their evaluations on students' actual academic performance (*meritocracy-as-mode*). However, besides this meritocratic criterion, the dominant script may gradually contain three other types of information: i) an automatic consideration of students' backgrounds (*habitus-as-mode*), ii) a more rational consideration of students' backgrounds (*habitus-rc-mode*), and iii) a rational consideration of additional criteria of students' academic success apart from their actual performance (*meritocracy-rc-mode*). Our main point is that on the individual level there is always one dominant script, but according to our multidimensional *and* gradual explanation of the emergence of teachers' evaluations, the conditions under which these evaluations are shaped may vary. Hence, according to the MFS, teachers' script selection can be understood as a belief-mediated mechanism in order to explain different determinants of teachers' evaluations.[63]

[63]In other contexts, one could object that the concept of habitus might also be understood as a *desire-mediated* mechanism: e.g. in sense of the desire for certain aesthetic standards (but also see Stigler

Results The question in the *CHiSP* data asking teachers to evaluate their students' prospective academic aptitude was phrased open-ended. Thus, teachers could either report a certain student to be able for academic studies, or to be not able from her point of view. This also means that we are confronted with a third category of students who were neither reported in one of the two other categories. Before the more complex structural model was estimated, we had to clarify whether this 'neutral' category could be regarded as an implicit 'middle' category – which was done via a conventional multinomial regression comparing the effect sizes of the predictors of interest for each of the two possible pairs. Since results indicated that the parameter estimates for the contrast *able vs. not able* are indeed remarkably larger than for the contrast *able vs. not mentioned*, we finally computed an ordered outcome variable with three categories (1 'not able'; 2 'not mentioned'; 3 'able').

In the structural model estimated based on a matrix of polychoric correlations (Olsson, 1979; Muthén, 1984; Aish and Jöreskog, 1990; Jöreskog, 1994) to account for the categorical measurement level of both our predictor variables and the outcome, we find significant effects on teachers' evaluations for each of the predictors deduced from the frame-selection model proposed above. However, effects for the indicators that had been linked to the idea of meritocracy (intelligence and average grade) are by far larger than for students' social backgrounds – which indicates that teachers' evaluations are largely accurate. Nonetheless, perfect accuracy would be realized only if the effects of students' social backgrounds were partialed out once analyses provide comprehensive controls for students' ability and achievement. One could argue that this is not the case in the model below, and that the teachers dispose of private information reflecting actual achievement and/or ability differences that correlate with the social background variables in our model. However, in survey data, comprehensive control for all possible indicators of a concept is hardly possible, which is why the alternative interpretation of remaining social background effects even having controlled for measures of both students' ability and achievement as a hint of residual *inaccuracy* of teachers' evaluations cannot be rejected, too.

One interesting result of our structural model deserving attention is that cognizant of all concerns for post-hoc model fitting, we could improve our model by endorsing separate regression paths for the analogy sub-score of intelligence[64] on both average grade and teachers' evaluations. This could either reflect a particular form of *meritocracy-rc-mode* processing in a sense that teachers consider students' aptitude for drawing analogy-

and Becker, 1977), or even for the (absolute) value of education per se. Indeed, cultural capital theory expects students' aspirations to vary among the social strata (Bourdieu, 1986; Bourdieu and Passeron, 1990; also see Dumais, 2006). This is another major difficulty of Bourdieu's theory as the assumption of class-invariant 'absolute' or 'idealistic' aspirations (Keller and Zavalloni, 1964; Boudon, 1974; but also see Stocké, 2011) would be violated. However, in our frame-theoretical account presented above, we understand students' knowledge of the symbolic codes demanded in the classroom as one anchor potentially shaping teachers' evaluations in terms of their *beliefs* about their students – which is why both the assumption of invariant desires and the interpretation of the impact of *habitus* as a belief-mediated mechanism can be defended.

[64] Intelligence was measured by an intelligence structure test (Amthauer, 1957) consisting of four different dimensions: i) analogy test, ii) word test, iii) number test, and iv) cube test.

based inferences as particularly relevant to arrive at well-founded both current (average grade) and prospective evaluations. However, as we argue in our conclusion, due to measurement error, it could well be that distinct effects of the verbal dimension of intelligence on both teachers' evaluations and students' average grade – being itself an aggregate of various teacher assessments – could imply a dimension of students' habitus that is not reflected in one of the indicators used in our structural model. In the outlook of this introduction (section 5), I again discuss the proposition made in the paper how this ambiguity could be addressed empirically.

4.2 Reference-Group Effects

The second paper is an example of applied reference-group research – but in contrast to what is conveniently known under the label of *Big Fish Little Pond Effects* (*BFLPE*; Marsh and Parker, 1984; Trautwein and Lüdtke, 2005; Dai and Rinn, 2008; Dijkstra et al., 2008; Marsh et al., 2008) or *Reflected-Glory Effects* (*RGE*; Cialdini et al., 1976; Marsh, 1987), respectively, the study in the volume at hand does not aim to explain students' self-concept as an outcome. Instead, the focus here is, as in the study previously described, on 10^{th} class teachers' evaluations. In the theoretical section of the paper, we deduce both class-average achievement and socioeconomic status as potential contextual-level determinants of teachers' evaluations, and we hypothesize that their effects – whatever sign their coefficient may obtain (which will be clarified below) – could vary by individual student achievement or teachers' grading concepts.

Since most of the research on reference-group effects in general and on *BFLPE* and *RGE* in particular has been invested by social psychologists, little effort has been made to link the respective theoretical assumptions to the underlying social mechanisms. However, classic multilevel theory provides a useful starting point for this endeavor.

In general, while multilevel *techniques* have made a lot of progress during the last decades (Bryk and Raudenbush, 1992; Goldstein, 1995; Snijders and Bosker, 1999; Gelman and Hill, 2007; Hox, 2010), I do not see that many of the critical *theoretical* remarks that have been opposed towards contextual effect research already beginning from the 1970s were addressed with similar effort.

Beginning with Davis (1966), reference-group effect research broadened the focus of educational sociology and its explanations of students' educational outcomes. While a brief summary of this and the following studies is given in section 2.1 of the paper, the crucial point to be made here is that about at the same time, a critical discussion emerged about how contextual-level effects can be thought to operate in terms of underlying mechanisms (Hauser, 1970a; Barton, 1970; Hauser, 1970b; Farkas, 1974; Hauser, 1974; Blalock, 1984; Van den Eeden, 1992). Following what had been called *ecological fallacy* (Robinson, 1950) regarding pure macro-level models, Hauser (1970a) introduced the term "contextual fallacy" in order to address theoretical identification problems in testing macro-to-micro hypotheses:

> "The contextual fallacy occurs when residual differences among a set of social groups, which remain after the effects of one or more individual attributes

> have been partialed out, are interpreted in terms of social or psychological mechanisms correlated with group levels of one of the individual attributes" (Hauser, 1970a, p. 659).

Precisely, aforehand to carrying out applied contextual-level analyses, it is of utmost importance (and probably even more important than in 'straightforward' individual-level models) that "[t]he exact meaning of 'group effect' needs specification in each research situation" (Hauser, 1970a, p. 661). Otherwise, the researcher can never rule out that individual (e.g. student-level) composition effects could be responsible for the suspected contextual-level effect. Therefore, after it has been ensured that a complete and correct individual-level model that is satisfying in both theoretical and empirical terms has been formulated, specification of convincing macro-to-micro mechanisms is inevitable:

> "Indeed, the greatest weakness of contextual analysis is the vagueness with which its causal mechanisms are usually specified. This vagueness seems to be the reason that many of its proponents slip into the contextual fallacy" (Farkas, 1974, p. 357).

Since the interpretation of contextual-level effects stands at risk of invalid social psychological assumptions for the former may represent a variety of mechanisms (Hauser, 1974; Blalock, 1984), major concern of contextual effect research in educational settings should be to isolate effects of, say, an intellectual climate in the classroom that is not mediated by or represents something more than interpersonal processes among students (Rigsby and McDill, 1972, p. 315), and that is "susceptible to unambiguous and distinctively sociological interpretation" (Hauser, 1974, p. 369).

Apart from what is usually denoted as 'main effects' (although this term is somehow misleading; see Friedrich, 1982; Brambor et al., 2006), special concern should also be devoted to cross-level interaction effects since already Kendall (1951, 189) noted that "[s]ignificant interactions do not necessarily imply interaction in any real sense. They may arise from heterogeneity in the data". Therefore, it is even more important to control for individual-level heterogeneity as much as possible before interpreting interaction effects.

Regarding reference-group effect research analyzing the impact of class-level achievement on students' self-concept, it has already been noted by Blalock (1984) that (negative) frog-pond effects (i.e. the *BFLPE*) and positive normative or climate effects (i.e. the *RGE*) may be additive, and in that case, they could not be separated. However, building on the preceding considerations, given a more elaborate *understanding* (in terms of social mechanisms) of the comparison processes involved, and given an adequate operationalization of indicators on both levels, it should be possible to approximate the social situation in the classroom.

A useful starting point for a mechanism-based explanation of contextual effects could again be the *DBO* model by Hedström (2005).[65] In that framework, it should consequently be the opportunity structure that forms the crucial link between social context

[65] For sure, other theoretical approaches to social mechanisms would be equally suitable to explicitly

in the classroom and social action on the individual level. Concretely, interaction should be understood as a form of social contact with a 'generalized other' (cf. Hedström, 2005, p. 44) that may influence *ego*'s action as could every single-person *alter*.

The original *BFLP* hypothesis postulates that independently of a student's individual ability, she will dispose of a worse academic self-concept when she finds herself in a high-ability class compared to being surrounded by classmates with lower ability on average. While the explanation suggested for this phenomenon – which yielded a great deal of supporting studies (Marsh and Parker, 1984; Marsh, 1987; McFarland and Buehler, 1995; Zeidner and Schleyer, 1998; Lüdtke et al., 2005; Rindermann and Heller, 2005; Trautwein et al., 2006; Seaton et al., 2010; Thijs et al., 2010) – usually refers to *contrast effects* resulting from upward comparisons (Collins, 1996), the implied *social mechanism* is one of relative deprivation (Davis, 1966): In *DBO* terminology, the opportunity structure in the classroom in terms of average ability alters students' beliefs in terms of their academic self-concept. Importantly, we should not assume this to be the result of "social telepathy", but of an *endogenous feedback structure* among students (Erbring and Young, 1979).

On the other hand, some studies found empirical support for a *RGE* that is defined as a *positive* impact of school- or class-level achievement on students' academic self-concept. While in the name-giving study by Cialdini et al. (1976), the study focus lay on players' identification with more successful football teams, simply denoting the empirical observation found in educational settings by the term 'assimilation' effect may be criticized to be a theoretical underspecification – particularly since school status was also found to exert positive effects on students' *aspirations*.

The latter effect may be reconstructed as a mechanism of *rational imitation* (Hedström, 2005, p. 49): In high status and/or ability environments, it is reasonable for a student to upward-adjust both aspirations and educational effort *since otherwise*, effects of deprivation would be even more disappointing.[66] A positive effect of school status or school achievement on students' academic self-concept could then be the result of realistic assessments regarding the increase in effort exerted (even apart from student GPA, as this variable is occasionally controlled for).

Although Marsh et al. (2000) were able to empirically juxtapose a *positive* effect of school status and a *negative* effect of class-level achievement on student academic self-concept, respectively, the *theoretical* question why both effects should be expected (and can be observed) to operate simultaneously still remains unresolved in *BFLPE* research. Clarification in that respect is provided in terms of the analytical distinction proposed by Manski (1993, p. 31; orig. emph.) who separates four different types of contextual-level effects (in a broader sense) from each other:

incorporate contextual effects. Particularly, Mario Bunge's integration of social mechanisms into his more abstract theory of *systemism* aims to bridge the gap between pure macro-level (i.e. *holistic)* and pure micro-level (i.e. *individualistic*) explanations (Bunge, 1997, 2004). However, since the *DBO* model is more close – though not necessarily glued – to rational-choice or subjective-expected-utility related action theories, it is the preferred conceptual framework also for the second paper.

[66] This mechanism corresponds to one possible explanation for the self-fulfilling prophecy to come true (Biggs, 2009, p. 309).

> "**Endogenous effects**, wherein the propensity of an individual to behave in some way varies with the prevalence of that behavior in some reference group.
> **Contextual effects**, wherein the propensity of an individual to behave in some way varies with the distribution of background characteristics in the reference group.
> **Ecological effects**, wherein individuals in the same reference group tend to behave similarly because they face similar institutional environments.
> **Correlated individual effects**, wherein individuals in the same reference group tend to behave similarly because they have similar individual characteristics."

While the term "endogenous effects" is used to refer to endogenous feedback effects in the sense of Erbring and Young (1979), contextual and ecological effects are *exogenous* in terms of a distinct impact on individual behavior apart from individual-level characteristics or interaction. The four different types can be illustrated by the example of school achievement effects (Manski, 1993, p. 31f.; emph. added):

> "There is an *endogenous* effect if, all else being equal, individual achievement tends to vary with the average achievement of the students in the youth's high school or ethnic group, or in another reference group. There is a *contextual* effect if achievement tends to vary with, say, the socioeconomic composition of the reference group. There is an *ecological* effect if students in the same school tend to achieve similarly because they are taught by the same teachers. There are *correlated individual* effects if students in the same school tend to have similar family backgrounds and these background characteristics tend to affect achievement."

Now it becomes understandable why the empirically observable negative *BFLPE* is one *net of* the *RGE*, and how it is analytically possible that both effects may operate simultaneously: The negative *BFLPE* is an *endogenous feedback effect* as a result of between-student interaction – inducing the belief-mediated mechanism of relative deprivation. The positive *RGE*, contrarily, is an *exogenous contextual effect* in terms of a school's student composition – inducing the belief-mediated mechanism of rational imitation.

But how can these considerations be related to teachers' evaluations? To be sure, naïvely transferring exactly the same mechanisms on a different outcome, even if the contextual-level variables are the same, would lead us astray. Instead, in case of (simple) reference-group (main) effects, we hypothesize an opportunity-mediated mechanism since it is the social situation in the classroom that is supposed to mediate teachers' evaluation practices.

Similarly to conventional *BFLPE* and *RGE* research, in the paper we use evidence from preceding analyses to deduce 'main effect' hypotheses opposing to each other: A *negative* effect of class-level achievement on a teacher's evaluation net of all individual-level covariates could be expected if she simply adjusts her reference standards according to the achievement level she observes in the classroom. In that case, teachers' beliefs and desires would remain constant while a 'pure' opportunity-mediated mechanism directly alters teachers' action in terms of assigning a specific evaluation to a student.

Contrarily, we also find arguments supporting a *positive* relationship between class-level achievement and teachers' evaluations: First, teachers' might adjust their evaluations according to a mechanism of *regression to the mean* (Galton, 1886; also see Healy and Goldstein, 1978) – which would also be a pure opportunity-based mechanism.[67] Furthermore, this effect might be amplified by a *concatenation* of an opportunity- and a belief-mediated mechanism: For instance, the belief-mediated mechanism of *dissonance reduction* could come into play in that teachers follow an implicit decision rule of the type, "a member of that bright class can't be that dull" (and reversely). This is what is conveniently denoted as a *Halo effect* (Thorndike, 1920).[68] Similarly, a comparable belief-mediated mechanism can be expected to hold for the implied effect of class-level socioeconomic status on teachers' evaluations in that teachers suspect parents in high-SES classes to be equipped with comparably higher educational aspirations which they also project onto students with comparably lower social backgrounds.

Having outlined the social mechanisms of our 'main effect' hypotheses, the former, however, are silent about teachers' action scripts as they have been described in the preceding section.

While one can assume that simple reference-group effects do not affect the operating mode of teachers' action scripts, this is not the case in the event of interaction effects. Given that student-level heterogeneity has been sufficiently controlled for (cf. Kendall, 1951), we hypothesize a negative interaction effect between class-level achievement and student achievement both as predictors of teachers' evaluations.

As commonly known, an interaction effect of the type $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2$ can be interpreted in two substantial ways: as an effect of x_1 on y varying with levels of x_2, or as an effect of x_2 on y varying with levels of x_1 (Friedrich, 1982; Brambor et al., 2006). For the interaction effect specified above, on the one hand, it is possible to assume the contextual-level effect of class-level achievement on teachers' evaluations to vary with levels of student ability – but on the other hand, it would be equally imaginable to hypothesize the effect of student-level achievement on teachers' evaluations to vary with levels of class-level achievement.

Providing a mechanism-based explanation for both readings, concerning the former, the 'pure' opportunity-mediated mechanisms of reference-standard adjustment (in case of a negative contextual-level effect) or regression to the mean (in case of a positive

[67] I should loose some more words on this. While Galton (1886) by origin referred to changes in a particular measurement from one generation to the next (such as body size), the psychological literature is usually talking about the fact that an observation with an extreme value in the first measurement can be expected to show a less extreme value in the second measurement (Tversky and Kahneman, 1974; Healy and Goldstein, 1978). Here, however, the idea is that teachers adjust their reference standards according to an *observable* mean (i.e. class-average achievement).

[68] Whereas the mechanism of dissonance reduction is usually applied to in-group member selection (e.g. Rigsby and McDill, 1972), its application onto the Halo effect is straightforward: One standard interpretation of the latter is the one of a discrepancy between a rater's *observed* and the *true* correlation of two characteristics (Murphy et al., 1993), in our case between classroom achievement – or social status – and a student's achievement nested in the respective classroom. This (positive) difference is due to a mechanism of dissonance reduction in order to minimize cognitive transaction costs in terms of unbalanced cognitive structures (Heider, 1946; Festinger, 1957).

contextual-level effect; see Galton, 1886; Healy and Goldstein, 1978) taken alone could be responsible for lower-achievement students to be more affected by the respective effect since in both cases, the high-achievers would serve as the foil all lower groups are compared to.[69]

Regarding the latter, a mechanism-based reconstruction can easily be incorporated in the frame-selection approach as outlined in the preceding subsection: If an attribute is all too common, it might forfeit its nimbus of distinctiveness. Hence, teachers in high-achievement environments could tend to leave their *meritocracy* script routine (in whatever mode operating) and switch to alternative (e.g. status-based) criteria in order to arrive at an evaluation. Hence, the opportunity-mediated mechanism of a reference-group effect would itself mediate the belief-mediated mechanism of script selection at the teacher level.

Finally, we hypothesize that the relationship between class-level achievement and teachers' evaluations varies by teachers' reported grade concepts. Concretely, we have in mind that teachers who report to follow more individualist grading concepts might be less affectable for reference standard adjustments due to contextual-level effects in terms of class-level achievement than teachers who report more relational grading concepts. In terms of social mechanisms, an opportunity-based mechanism would in turn be mediated by teachers' beliefs.

Results By means of a series of cross-classified multilevel models wherein the outcome of teachers' evaluations[70] is nested in both student and teacher contexts with school classes as the highest unit of analysis, we found that in contrast to both conventional *BFLPE* research and reference-group studies with German primary school teachers' transition recommendations as an outcome, teachers' evaluations are positively affected by both class-mean intelligence and class-mean grade point average (GPA). This is interpreted in sense of a *Halo* effect (Thorndike, 1920) – which I have just reconstructed as a concatenation of an opportunity- and a belief-mediated mechanism. To be sure, the alternative interpretation of a *regression to the mean* (Galton, 1886; Healy and Goldstein, 1978) – a pure opportunity-based mechanism – cannot be rejected either. The particular social composition of our sample (*Gymnasium*, i.e. German highest-track secondary school students only), the lower extent of obligation associated with non-binding, anonymous assessments, and their comparably later point in time of measurement may explain why our results also differ from reference-group effect studies based on primary school teachers' recommendations. Still, model fit criteria (in terms of pseudo R-squared) reveal that individual-level achievement is a much more important predictor of teachers' evaluations than respective contextual-level indicators.

[69] It has been observed that halo effects seem to be stronger for ratees less familiar to the rater (Kozlowski and Kirsch, 1987). Low-achievement students might be less noticeable in the classroom due to their lesser class contribution.

[70] Here we only analyzed whether a teacher assessed her student to be 'able' vs. to be 'not able'. For a multilevel analysis of the contrast 'able' vs. 'not mentioned' based on the same data see Becker and Birkelbach (2010).

While a significant interaction effect between class-average and student-level intelligence could not be detected, it was noted that the effect of class-average GPA on teachers' evaluations significantly increases with better student marks. This contradicts our hypothesis which expected this effect to be negative. Also, comparing the marginal effects of both readings of this interaction term suggests that the effect of class-level GPA on teachers' evaluations does vary more strongly with student-level GPA than does the effect of student-level GPA with class-level GPA. Hence, results suggest that the significant interaction term is indeed due to an opportunity-mediated mechanism that is itself mediated by a belief-mediated mechanism on the teacher level rather than due to a belief-mediated mechanism of script selection mediated by the opportunity-mediated mechanism of a reference-group effect.

What we did not find is statistically solid evidence of class-average achievement effects to vary by teachers' grading concepts. Although simple graphical insight in the respective marginal effects tends in the direction of *weaker* contextual-level effects for teachers with more *individualist* grading concepts, and *stronger* effects for teachers with more *relational* grading concepts – which would challenge both our theoretical explanation given in terms of a concatenation of an opportunity- and a belief-mediated mechanism as well as assumptions in already existing studies (Rheinberg, 1980; McFarland and Buehler, 1995; Marsh et al., 2001; Lüdtke et al., 2005; Seaton et al., 2009) – we caution against overhasty accepting this result since conventional statistical significance levels were not reached.

4.3 A Formal Model of Self-Fulfilling Teacher Expectancy Effects

The idea of a self-fulfilling prophecy can be said to be one of the most prominent examples of a social mechanism. The original idea is ascribed to Robert Merton (1948)[71] who illustrates how a potentially mislead rumor of a bank's illiquidity might nonetheless cause the bank's bankruptcy: Alienated by the rumor, the first customers will withdraw their savings – which will in a second step move other costumers to follow suit. In the end, panicky withdrawals might in fact lead to the bank's breakdown – although the initial rumor did not necessarily correspond to the bank's initial financial situation.

This seminal description of a self-fulfilling prophecy was convincingly reconstructed by Hedström (2005, p. 48) as a *belief-mediated* mechanism: The beliefs of the first depositors who withdraw their savings affect the beliefs of the remainders who now have good reasons to assume that there might actually be something wrong with the bank and thus also withdraw their savings. Or phrased more analytically (Biggs, 2009, p.

[71] Actually, before also Popper (1944a, p. 89) referred to "[t]he idea that predictions may influence predicted events". In the monograph *The Poverty of Historicism*, he labeled this idea "Oedipus effect" (Popper, 1957, p. 11). However, Popper might have got familiar with the idea by Merton's (1936) paper wherein he just briefly sketched the idea of how a prediction might change a course of development by altering an actor's social situation (also see Birkelbach, 2011, p. 134). Even earlier, Thomas and Thomas (1928, p. 572) accurately verbalized, "If men define situations as real, they are real in their consequences" – a statement well-known as the *Thomas Theorem* which Merton (1948, p. 193) also discusses at the beginning of his more famous paper.

295):

(1) X beliefs that 'Y is p'.

(2) X therefore does b.

(3) Because of (2), Y becomes p.

The above self-fulfilling prophecy should be distinguished from the inductively-derived prophecy which can be reconstructed as follows (Biggs, 2009, p. 296):

(0) Y is p.

(1) Because of (0), X beliefs that 'Y is p'.

(2) X therefore does b.

(3) Because of (0), Y manifests p.

In the first case, it is an actor's inaccurate *belief* about the social situation that makes the difference, while in the second case, it is the social situation that causes an accurate belief – which can be illustrated by imposing counterfactuals on the respective first condition: If in the first case, X were to belief that 'Y is q', X would do c instead of b and thereby cause Y to be p – while in the second case, if Y really were p, X falsely believing Y to be q would not make a difference (Biggs, 2009, p. 296).[72]

In educational sciences, the notion of a self-fulfilling prophecy is conveniently used to refer to what has been labeled the *Pygmalion effect*. In 1968, Rosenthal and Jacobson (1968) published an influential study named *Pygmalion in the Classroom*. In a quasi-experimental design, the authors first administered a non-verbal intelligence test to elementary school children. Towards the teachers, the authors pretended that their study was aimed at the identification of so-called "late bloomers" – students who can be expected to show a sudden intellectual spurt in course of the upcoming term. However, when communicating students' test results to the teachers, the authors named a randomly-chosen set of students to be the "late bloomers" – which had nothing to do with their actual test results. In a follow-up achievement test one year after, though, the artificially-created group of late bloomers scored significantly higher than the control group. Hence, the authors concluded that just as the Cypriot sculptor Pygmalion in Ovid's *Metamorphoses* created a statue so beautiful that he fell in love with it, and due to his caress, invoked it to life, a particular teacher treatment effect invoked by a plus of (false) information may yield actual changes in student behavior and related outcomes.

[72] The self-fulfilling prophecy should also be distinguished from the *Matthew-Effect* of cumulative advantage that operates via an *opportunity*-mediated mechanism (DiPrete and Eirich, 2006; also see Birkelbach, 2011, p. 137).

Theoretical identification of SFP's While the initial *Pygmalion in the Classroom* was set up in a quasi-experimental design, most of what followed in teacher expectancy effects research relied on survey data (see overviews by Jussim, 1986; Jussim et al., 1996; Jussim and Harber, 2005). Regarding isolation of a causal effect, survey data are always inferior to a true experimental design (Biggs, 2009) – though econometricians have developed sophisticated methods to get rid of a great deal of heterogeneity in the data that would not be equally problematic in an experimental setting (Gangl, 2010, also see below). What survey research *can* do, though, is to approximate the social mechanisms that stand behind the causal effect – leading to a better *understanding* of the phenomenon to be isolated by means of methods. In other words, a crucial prerequisite of empirical identification of a teacher treatment effect is a thoroughly-specified theoretical model with preferably fine-grained bridge assumptions.

To be precise, survey data research faces both the burden and the benefit of a particular specification problem – which in the case of self-fulfilling prophecies translates into the following questions: 1) To what extent are teachers' initial beliefs about their students accurate; 2) how are initial teacher perceptions affected by subsequent student behavior; and 3) in what manner do *inaccurate* teacher perceptions affect students' educational outcomes? The beneficial part of that requirement lies in the chance to get closer to the social mechanisms that operate beneath the surface of a self-fulfilling prophecy. So let's now treat all three questions systematically.

To what extent are teachers' initial beliefs about their students accurate? A major objection against self-fulfilling prophecy research from the very beginning is the argument that teacher expectancy effects on student achievement reflect *accurate* beliefs of teachers that correspond to unobserved student characteristics. Jussim (1986) distinguishes between a strong and a weak form of the accuracy argument. The strong form of the argument claims that teachers' expectations based on stereotyping will always be inaccurate, but teachers' expectations based on direct observation of their students will always and necessarily be accurate. In its weak form, the argument does not make the claim of absoluteness but only postulates that direct observation of student behavior will be *more* accurate than expectancy formation based on stereotyping.

Following Jussim (1986, p. 431), it is hard to test the accuracy of teachers' expectations *just because* they may invoke self-fulfilling prophecy effects. However, as Jussim et al. (1996, p. 288) summarize the existing research up to that point, a large degree of teacher expectancy effects can be attributed to the fact *that* they are accurate – a conclusion that is maintained in a more recent review by Jussim and Harber (2005).[73] However, a current study by Ready and Wright (2011) could show that holding between-group achievement differences constant, teachers are especially error-prone in lower socioeconomic and lower achievement classroom contexts. Regarding teachers' expectations as measured by teachers' evaluations in the data at hand, the first two papers in this volume demonstrate their vulnerability from both social background and reference-group effects apart from students' intelligence and grade point average – which might be a hint of inaccuracy to some extent.

[73]"Prediction without causation is exactly how we define accuracy" (Jussim and Harber, 2005, p. 141).

How are initial teacher perceptions affected by subsequent student behavior? Jussim (1986) summarizes evidence that teachers may maintain their expectations even though they are biased and teachers are confronted with contradicting information. In contrast to teachers with more flexible expectations, it's the teachers with rigid expectations that are suspected to generate self-fulfilling prophecy effects since the former can be expected to change their opinion when confronted with disconfirming evidence. A social mechanism that may explain rigid expectations is the belief-mediated mechanism of dissonance theory (Jussim, 1986, p. 430): When expectancy-consistent information is more likely to be remembered, evidence about a good performance of a presumably bad student has less chance to change teachers' beliefs than a good performance of a presumably good student to maintain teachers' beliefs. Moreover, teachers may perceive students' behavior as consistent with their expectations because of *perceptual biases* (Jussim et al., 1996, p. 286).[74]

In what manner do inaccurate teacher perceptions affect students' educational outcomes? Unsurprisingly, the answer to that question is closely related to the argument of accuracy. Only if *inaccurate* teacher expectations can be identified, it can be ruled out that teacher perceptions are valid predictions of unobserved characteristics that actually cause student achievement.

But even if inaccurate teacher expectations can be identified empirically (see below), it has still to be clarified *how* different teacher expectations cause different student outcomes. As Jussim (1986, p. 435) notes, teacher treatment effects are supposed to occur when teachers provide more emotional support, clearer and more favorable feedback to high expectancy students, pay more attention, and teach more (and also more difficult) material to the highs.[75] While most (if not all) self-fulfilling prophecy studies tested for teacher expectancy effects of subsequent student achievement, in the paper, I develop a theoretical model to explain how teachers' expectations may affect students' *educational transitions*. It is beyond the scope of this introduction to replicate the mathematics again, but the model's crucial idea should be elaborated on notwithstanding: As described above, the model of relative risk aversion by Breen and Goldthorpe (1997) and a conceptual similar model by Esser (1999) explain students' educational transitions by a multiple parameter model covering both subjective costs as well as expected benefits of education. In both models – as well as in a theoretical predecessor proposed by Erikson and Jonsson (1996a) –, the subjective expected probability of educational success is a crucial parameter and is assumed to vary with preceding educational performance

[74] While a self-fulfilling prophecy is defined as the effect of a teacher's inaccurate expectation on student achievement, and rigid teacher expectations may emerge because dissonant information is ignored, I wish to argue here that perceptual biases can also be reconstructed by a mechanism of dissonance reduction: In a trivariate relation between *ego* (a teacher), *alter* (a student), and a *state-of-affairs* (e.g. school performance), it would be rational for *ego* to form beliefs of *alter*'s behavior consistent with *ego*'s preceding beliefs in order to avoid cognitive dissonance – regardless of whether the beliefs actual correspond to *alter*'s behavior. For teachers, it might be efficient to ignore 'dissonant' student behavior due to a strategy of satisficing (Simon, 1957).

[75] Following (Biggs, 2009, p. 308f.), teachers providing more emotional support would affect students' *beliefs*, while teachers providing more challenging material to higher-achievers would alter their *opportunity structure*.

(Breen and Goldthorpe, 1997, 285; Esser, 1999, p. 272).

Recalling what has been said about the impact of a teacher treatment effect, the efficacy of a self-fulfilling prophecy can be modeled in terms of a student's subjective expected probability of educational success which is a function of preceding expected educational success, actual performance, and a teacher treatment component consisting of factors such as classroom praise, bilateral encouragement, etc. Moreover, I follow the literature by assuming that inaccurate teacher expectations can be defined as teachers' over- and underestimations of their students compared to a set of student background variables (also see below).

If all other model parameters are restricted to be constant, an overestimated student would expect a higher utility from continuing on a higher educational track than an underestimated student due to her higher subjective expected probability of educational success. This can be attributed to actual achievement differences on the one hand, and residual differences in motivation and self-concept (cf. Jussim, 1989; Gill and Reynolds, 1999; Muller et al., 1999; Mechtenberg, 2009; Mistry et al., 2009) on the other hand.

By relating the effect of a self-fulfilling prophecy to differences in the subjective expected probabilities of educational success of over- and underestimated students, respectively, the underlying teacher treatment effect is reconstructed as a *concatenation of belief-mediated mechanisms*: As outlined above, teachers first form initial beliefs about a student's academic performance, and although these beliefs are largely accurate, a residual share of inaccuracy was shown to remain (Ready and Wright, 2011). By a mechanism of dissonance reduction, teachers might ignore positive information about negative students (and reversely) and thereby form rigid expectations that are prone to induce a self-fulfilling prophecy. Thus, the latter can be defined as the impact of inaccurate teachers' expectations on students' beliefs in terms of their subjective expected probability of educational success – caused by both achievement differences and direct effects of teacher treatment on students' motivation and self-concept – inducing differences in the utility that is attached to a higher educational outcome between over- and underestimated students.[76] Hence, by linking a convenient SEU explanation of social inequality in educational opportunities (Esser, 1999) to self-fulfilling prophecy research dealing with inaccurate teacher beliefs, we now see the importance of Boudon's demand for an action theory that has to deal with evidently false beliefs (Boudon, 1996, 1998, 2003).

Empirical identification of SFP's As just described, it is crucial to arrive at an empirical identification of *inaccurate* teachers' expectations since otherwise, it cannot be ruled out that teachers predict rather than cause subsequent student achievement. In contemporary self-fulfilling prophecy research, this is accomplished by regressing the

[76]While all cost-benefit models of educational transitions assume that desires (i.e. the absolute value of education) are constant between the social strata, it might well be possible that additional opportunity-mediated mechanisms – e.g. of classroom context – affect the utilities attached to educational transitions in general and the effect of a self-fulfilling prophecy as described above in particular (Blalock and Wilken, 1979). Though certainly promising, this is beyond the scope of both the underlying paper and this section.

available measure of teachers' expectations on a set of student background variables and by then storing the residuals of that regression as teachers' over- and underestimations *compared to* the respective background characteristics. As predictors, both student achievement and motivation are usually considered to isolate a residual component in a teacher's expectation that might reflect inaccuracy (Madon et al., 1997, 2006).

Unfortunately, with the *CHiSP* data, I am not able to operationalize indicators for all parameters in the theoretical model at each point in time. The data entails measures for student cost-benefit assessments and teachers' expectations measured at the same time – which might be problematic for the empirical identification of self-fulfilling prophecies. However, I assume that by regressing teachers' evaluations – the measure of teachers' expectations in the *CHiSP* data – on suitable indicators of students' performance and motivation, respectively, and by isolating the residual terms resulting from these regressions, differences in students' subjective expected success estimates that are not reflected in actual performance or motivation differences can at least be approximated. Once a significant effect of the residual terms on students' educational transitions is found – as it holds for students' probability of passing German Abitur – following good falsificationist practices (Popper, 1959), it is up to future studies to reject the assumption that this effect is not due to a belief-mediated mechanism of a teacher treatment effect.

The statistical approaches in the paper are aimed at providing conservative (i.e. lower-bound) estimates of self-fulfilling prophecy effects (also see Biggs, 2009, p. 300) by also trying to control for unobserved heterogeneity. To be precise, it might still be the case that teachers dispose of something like *private information* (Cunha et al., 2005; Cunha and Heckman, 2007) that is not reflected in measures of either student motivation or performance, but affects their transition propensities (either directly, or indirectly via cost-benefit considerations[77]). This is, admittedly, an identification problem of social mechanisms based on survey data: "Under non-experimental conditions we can see only what that mechanism *in conjunction with other factors* makes it do" (Collier, 1994, p. 33). To avoid potential biases as accurately as possible, I apply two different methods of controlling for potential unobserved but confounding variables – a so-called *Heckman* model (Heckman, 1979) on the one hand, and a sensitivity analysis (Buis, 2007, 2010, 2011) on the other hand. Both approaches suggest that at least teacher treatment effects on students' probability of passing Abitur are not affected by unobserved heterogeneity (see below).

Results Analyses in this and the following paper differ from the two preceding ones in that teachers' evaluations are used to construct *predictors* rather than the outcome. In order to arrive at a satisfactory empirical identification of self-fulfilling prophecies, teachers' evaluations are regressed on measures of both student performance and motivation.

[77] If students from different social strata would have different *desires* with regard to the absolute value of education *per se* – which would contradict theoretical approaches in the tradition of Keller and Zavalloni (1964) as well as Boudon (1974) –, and if differences in these desires would have an impact on students' transition decisions, and if teachers could observe these differences and would ground their expectations upon them, this private information needs not necessarily be reflected in students' cost-benefit considerations.

Regarding performance, regressors concerned student intelligence as measured by the intelligence structure test (Amthauer, 1957) and student grade point average (GPA). As regards motivation, indicators of student homework effort, their self-assessment of relative school performance compared to classmates, and general self-confidence are considered. Results of these analyses show that indicators of performance are much more important predictors of teachers' evaluations than indicators of student motivation (whereupon the effect of student self-assessment of relative school performance is highest).

Separate residual terms for each of the performance model, the motivation model, and a full model comprising predictors of both concepts are stored in the data and used to predict i) students' probability to pass Abitur (i.e. German highest secondary-school track exam) on the first try, and ii) students' propensity to start academic studies within two years after high school graduation.[78] All models control for indicators of all subjective-expected-utility (SEU) concepts conventionally used to predict students' educational transitions. These comprise controls for the expected educational benefit, the subjective expected probability of educational success[79], the expected status decline, its expected influence on actual transition decisions as well as the expected costs of education. Additionally, the models account for potential *heterogeneity bias* that may arise due to unobserved variance of both the SEU indicators[80] and the residual terms used to approximate teacher treatment effects in terms of a self-fulfilling prophecy, as well as for *selection bias* following from the fact that we can only estimate university entrance transition probabilities for those students who successfully graduated from high school. This is accomplished by means of a so-called Heckman correction (Heckman, 1979) that has already been applied on educational transition models (Becker, 2000, 2003). Following this approach (and extending it to account for potential heterogeneity bias also among the indicators used to identify self-fulfilling prophecies), all cost-benefit indicators as well as the residuals obtained to identify differences in subjective expected probability estimates net of actual motivation and performance are regressed on parental social class. When the results of the respective probit regressions are stored as *Inverse Mill's Ratios* and then entered as control terms in the equation of primary interest (i.e. predicting students' transition probabilities), it is possible to control for class-related differences that might enter the educational utility function (for a description of the particular mathematics of this approach see e.g. Heckman, 1979; Winship and Mare,

[78] See sections 3.1 and 4.1 of the respective papers for more details on the cut-offs used for the dependent variables.

[79] Note that due to data constraints, I do not dispose of an appropriate measure of *students'* subjective expected probability of educational success – which would certainly be a better measure in this stage of students' educational life course (also see the paragraph on the *waning coefficients pattern* in section 2.2 of paper 4) – than *parental* assessments. However, the respective bridge assumption in this context is that parents have good knowledge of their children's probability estimates and that the extent of these approximations does not vary between the social strata. Also note that this measure is analytically distinct from the concept used for theoretical identification of self-fulfilling prophecies as outlined above.

[80] One could argue that the identification assumption that the degree of accuracy regarding parental assessments of their children's subjective expected probability of educational success must not vary by parental social class is relaxed in models with heterogeneity bias correction.

1992; Breen, 1996; Briggs, 2004). As an alternative, I also apply a procedure to control for unobserved heterogeneity which was proposed by Buis (2007, 2010, 2011): A random variable with zero mean and both varying standard deviation and varying correlation with the self-fulfilling prophecy residuals is used to approximate a potential mediator of these residuals that might not be associated with parental social class. The advantage of that latter method is that is does not rely on the identifying assumption that parental social class as the crucial predictor in the selection equation is not associated with the dependent variables in the outcome equations – which might be an objection against the models applying the Heckman correction method.[81]

In particular, I found that self-fulfilling prophecies in terms of a teacher treatment effect – that was identified by residualized teachers' expectations – have a distinct effect for all three variants of the estimated residual terms (performance model, motivation model, and a full model consisting of indicators for both of these concepts) on both students' probability of passing Abitur as well as on their propensity to start academic studies apart from all indicators for the SEU predictors conventionally used in educational transition research.

Regarding corrections for selection bias, results remains stable in case of high school graduation (Abitur) – but not in case of students' propensities to start academic studies. The first result indicates that the belief-mediated mechanism of a self-fulfilling prophecy in terms of a teacher treatment effect is not affected by heterogeneity in beliefs that could be accounted for in terms of social class. The second result is an illustration of path dependency effects of self-fulfilling prophecies, but it also indicates that there is no distinct belief-mediated mechanism operating that affects university transitions once homogeneity in beliefs conditional on having successfully graduated from high school has been taken into account.[82]

Regarding sensitivity analyses, private information of the teachers that might be associated with either student beliefs of educational success or their educational transitions which can not be accounted for by means of parental social class should only be an issue distorting the robustness of the results if an appropriate indicator would correlate unreasonably high with both the residual terms that were used for empirical identification of self-fulfilling prophecies as well as with students' educational transitions.

Hence, I conclude that the belief-mediated teacher treatment effect in terms of a self-fulfilling prophecy predicts students' high school graduations robustly regarding i) heterogeneity of student beliefs according to social class, and ii) reasonable degrees of teachers' private information. Contrarily, once homogeneity of student beliefs in terms of

[81] Note that Birkelbach (2011) also tests for self-fulfilling prophecies in terms of a belief-mediated mechanism based on *CHiSP* data. However, although referring to the SEU explanation of educational decisions (Esser, 1999), he does not provide a *formal model* in SEU terminology. Furthermore, (in my view) less precise indicators of self-fulfilling prophecies without controls for endogeneity or selection bias are used – which is why I still consider my contribution to be of distinct scientific relevance.

[82] Given the stability of the indicator used to operationalize students' subjective expected probability of educational success, one could argue that relaxing the assumption that the proximity of this (parental) measure to students' own beliefs must not vary by social class is permissible from an empirical point of view.

a selection bias – arising conditional on having successfully graduated from high school – has been controlled for, we do not find anything more than path dependency effects.

4.4 Changes of Self-Fulfilling Teacher Expectancy Effects Over Time

The fourth paper asks, *Does the effect of teachers' expectations on students' educational opportunities decrease over educational transitions? A statistical matching approach.* Conceptually, I use the same framework for identification of a teacher treatment effect in terms of a self-fulfilling prophecy as in the third paper. However, in addition to the latter article, I also test for potential *changes* of this effect over students' educational life course.

Evidence from educational transition research Starting with conventional IEO research, several theoretical approaches postulate decreasing social background effects over students' educational transitions. First, according to the *l*ife *c*ourse *p*erspective (LCP; Müller and Karle, 1993), the more students make their way through secondary socialization, the less strong parental treatment effects may influence students in their academic outcomes. Second, *m*aximally *m*aintained *i*nequality (MMI; Raftery and Hout, 1993) suggests that decreasing social background effects at later educational transitions are in turn sensitive to the progress of educational expansion that affects enrollment rates. Third, *r*elative *r*isk *a*version (RRA; Breen and Goldthorpe, 1997) assumes that homogeneity in ability among working-class students increases with subsequent transitions; and once a particular transition has been passed, the risk of status decline for working-class parents is lower than for the preceding decision since the critical level of status maintenance is already reached. Fourth, *e*ffectively *m*aintained *i*nequality (EMI; Lucas, 2001, 2009) additionally accounts for the fact that educational systems may be tracked. Lucas (2001) finds persisting residual class differences between different educational tracks also at a higher transition – which rejects LCP and also challenges the assumption of MMI that class differences in educational transition probabilities diminish once a certain level of education has become universal due to educational expansion.

Formal evaluation of all propositions As LCP could already be rejected since results are more in favor of MMI (Lucas, 2001), in his formal analysis, Lucas (2009) only compares the implications of MMI, RRA and EMI. As he notes, only one of the three proposals, namely RRA, has already been written down in formal terms. Thus, Lucas' objective is first to provide a formal notation of both MMR and EMI, and second, to test which of the three proposals suffer from the logical threats *tautology*, *self-contradiction* and *evaluative infeasibility*.

According to Lucas (2009), the statement of MMI that "transition rates and inequality (as measured by odds ratios) remain constant unless forced to change by increasing enrollments" (Raftery and Hout, 1993, p. 42) can be understood in two ways: First, transitions by class odds-ratios ω_1 are a function of the effect of social origins on education demand, β, and the *change* in this effect over time, γ:

$$\omega_1 = h(\beta + \gamma) \tag{2}$$

Concluding that unless forced, odds-ratios will not change, implies that $\gamma = 0$ (Lucas, 2009, p. 466). Second, however, *if* transition rates are forced to change by increasing enrollments, according to Lucas (2009), the only conceivable way that margin-free odds-ratios could be forced to change would have to read

$$\omega_2 = h((\beta + \gamma), (\alpha + \lambda)) \tag{3}$$

– where α denotes the effect of population size on educational demand, and λ again the change in this effect over time. Hence, a change in the effect of population size on educational demand could force to change the transition odds-ratios without any change in the origin effect.

By plotting the results of an analysis of the main MMI statement in a logistic regression framework, Lucas (2009) can show that i) odds-ratio ω_2 is not margin-contaminated, as MMI would require; and ii) ω_1 is not forced down if high-origin transition probabilities increase towards 1. Thus, MMI is either contradictory (in case of ω_2) or evaluatively infeasible (in case of ω_1).

Regarding RRA and EMI, both proposals stand the test of being non-tautological, non-contradictory and evaluatively feasible. As both RRA and EMI share a lot of common statements with yet a number of considerable differences, Lucas (2009) is unable to get to a final statement whether the two theories relate to each other in a complementary (or even nested?) manner.

What is common among all theoretical accounts presented above is that by origin, they are not aimed at explaining changes in background effects over educational transitions – but over student cohorts. This phenomenon has already been reconstructed as an opportunity-mediated social mechanism by Hedström (2005, p. 55): "Differences in social mobility rates between different nations or between different points in time are explained by reference to differences in mobility opportunities due to differences in occupational or class distributions." This is approximately true for MMI (though formally rejected) since population growth is supposed to affect individual transition ratios. However, reading EMI more closely reveals that what it actually postulates should be reconstructed more accurately as a concatenation of an opportunity- *and* a *desire-mediated mechanism* since MMI also supposes population size to affect *educational demand* (captured by parameter α in Lucas' formal evaluation).

Regarding RRA, its strong anchoring on students' *beliefs* was already carved out – which reflects that a simple opportunity-mediated mechanism would fall too short. Finally, one could conclude that just as MMI, also EMI challenges RRA's assumption of constant desires among the social strata since it postulates that classes differ in their placement on available educational tracks (academic vs. vocational) – which may thoroughly be due to class-specific differences in educational desires.

However, for the objective of the present study, the particular applicability of RRA stems from the fact that it explicitly links students' ability to their subjective expected probability of educational success (while EMI is agnostic on this relationship; see Lucas, 2009, p. 502). As this is the crucial link the formal model of self-fulfilling prophecies presented in the third paper points to, RRA might be more suitable for deducing waning-coefficient hypotheses in this regard.

Bayesian belief updating This is especially true since in later writings, Breen (1999) and Breen and García-Peñalosa (2002) propose the mechanism of *Bayesian learning* to account for student belief updating conditional on having passed an educational transition. In social mechanism theory, learning theory in general can be regarded "as a specific type of DBO theory that is applicable when actors use information about the past to decide what to do in the future " (Hedström, 2005, p. 41).

The core idea of the reasoning followed in the paper is that students update their beliefs while following the Bayesian rule, which is illustrated in figure 2 (cf. Lynch, 2007, p. 10f.). Let us say that $p(A|B)$ denotes the area consisting only of A while assuming that B is already true. (The '|' is read as 'given', so $p(A|B)$ means the probability of A given B.) If we know that B is true, then the total sample space from the entire rectangle is reduced to the circle B only. *Given* this reduced space B, $p(A)$ is in turn reduced to (A, B) – which is given by the $A \cap B$ region. If A and B were independent, $p(A, B) = p(A) \cdot p(B)$. Thus, knowing that B is already true, it follows that

$$p(A|B) = \frac{p(A, B)}{p(B)} \tag{4}$$

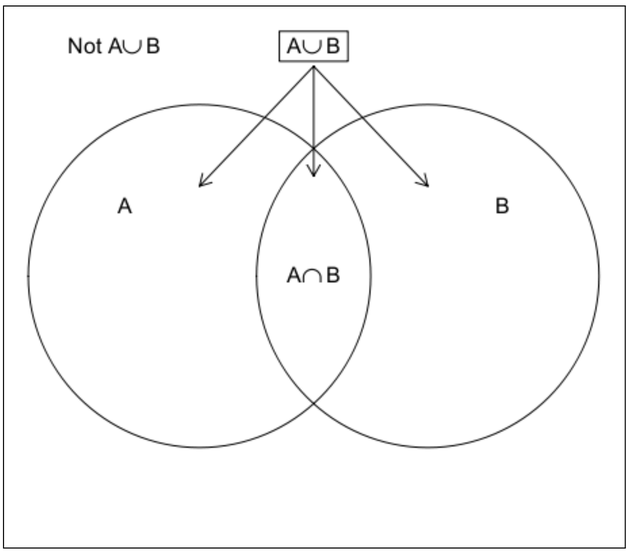


Figure 2: A Venn diagram illustrating the Bayesian rule. Source: Lynch (2007, p. 11).

Breen (1999) and Breen and García-Peñalosa (2002) proposed to model students' belief updating mechanisms – having passed a given transition – in a way of Bayesian learning that follows the above-shown rule. If θ' denotes a state wherein effort has a stronger impact on the probability of success than ability, and θ a state wherein it would be just the opposite, and AH a student's choice of a high educational career path, then the posterior belief of a student about θ' (i.e. the belief about θ' conditional on having succeeded in the high educational career path AH) reads (Breen and García-Peñalosa, 2002, p. 909):

$$Pr(\theta'|AH) = \frac{Pr(AH|\theta')Pr(\theta')}{Pr(AH|\theta')Pr(\theta') + Pr(AH|\theta)Pr(\theta)} \tag{5}$$

Hence, students' posterior belief about θ' is a function of their prior belief about θ' times the prior belief about their success in the higher academic track given that θ' is true (i.e. that effort is more important than ability) – divided by the term just described *plus* the prior belief that θ is true (i.e. ability is more important than effort) times the prior belief about their success in the higher academic track given that θ is true.

Grounding on these theoretical considerations, in the paper I use the mechanism of *Bayesian learning* proposed by Breen (1999) and Breen and García-Peñalosa (2002) to account for students' belief updating in course of their educational transitions. Admittedly, I do not present a comprehensive Bayesian model of educational transitions – which would go beyond the scope of a primarily empirical contribution. Notwithstanding this issue, the theoretical specification should hopefully become clear even though: Recall that according to what has been postulated regarding the theoretical specification of a teacher treatment effect on students' educational transitions in sense of a self-fulfilling prophecy, the subjective expected probability of educational success of a student who has been overestimated should be higher than the respective self-estimate of a student who had been underestimated by her teacher. Now the crucial assumption is that given an educational transition was successfully passed, this belief is updated upwardly. Due to this mechanism of belief updating conditional on having successfully passed the preceding transition, in the paper, it is illustratively sketched that variation in student beliefs should become more homogeneous – leading to decreasing self-fulfilling prophecy effects over time.

This rationale is then used to deduce testable hypotheses. Apart from the 'baseline' assumption (which is also tested in the third paper) that self-fulfilling prophecies in terms of a teacher treatment effect may have an impact on students' transition probabilities (which is here tested by means of different data; see below), two different hypotheses address potential changes of this effect over time: First, I postulate that the effect of a teacher's expectation which is measured before the first transition point *on the first transition* is larger than the effect of a teacher's expectation which is measured before the second transition point *on the second transition*. Second, I hypothesize that the long-term effect of a teacher's expectation measured before the first transition point *on the second transition* is smaller than the two possible short term effects. I expect both phenomena to occur due to a mechanism of Bayesian updating of student beliefs conditional on having successfully passed the first transition.

Results To evaluate all hypotheses that were deduced above, measures of teachers' expectations at two different points in time as well as two different transition points

– *idealiter* temporarily closely following upon such a measure – have to be observed. Unfortunately, the *CHiSP* only entails indicators of teachers' expectations measured in students' 10^{th} class, and nothing more. This is why other data sources have to be taken into account.

Since I did not find data that satisfies the above conditions and also entails accurate indicators of students' cost-benefit considerations in sense of Esser's (1999) SEU model of educational transitions, I make use of the technique of *statistical matching* (Rubin, 1986; D'Orazio et al., 2006) in order to arrive at an artificial longitudinal data file by means of linking distinct student data files according to a distance function.

Concretely, apart from *CHiSP*, I rely on a data set named "Elternhaus und Bildungschancen" (*P*arental *H*ome and *E*ducational *O*pportunities; henceforth abbreviated as *PHEO*) which was surveyed at about the same time (1968) as the former data. The measure of teachers' expectations available here are primary school teachers' transition recommendations regarding whether or not a given student should better attend German highest track school (Gymnasium) – or a lower form.[83] Consequently, the corresponding educational transition taken from *PHEO* are students' transitions from primary to secondary school (here also simplified to the question whether or not students made it to Gymnasium). Additionally, comparable variables in order to operationalize students' cost-benefit considerations as well as controls for parental social class could be derived. As in *CHiSP*, teachers' expectations are residualized by using appropriate indicators of students' performance and motivation, respectively, in order to arrive at an empirical identification of a self-fulfilling prophecy in terms of a teacher treatment effect by isolating the inaccurate part in teachers' expectations.[84] Both data are linked in order to arrive at an artificial longitudinal data set by means of *statistical matching*. Concretely, a vector of variables $\mathbf{X}$ from file A and another vector of variables $\mathbf{Z}$ from file B – with each of them missing in the other file, respectively – can be combined based on a vector of variables $\mathbf{Y}$ common to both files as long as $\mathbf{X}$ and $\mathbf{Z}$ are independent conditional on $\mathbf{Y}$. This *conditional independence assumption* cannot be tested empirically but only evaluated theoretically. As it is outlined in more detail in paper 4, I expect both teachers' evaluations and students' transition recommendations in primary ($\mathbf{X}$) and secondary school ($\mathbf{Z}$) to be independent of each other conditional on a vector of students' cost-benefit considerations $\mathbf{Y}$.

Having identified students' expected benefit, their expected amount of status decline and their subjective expected probability of educational success as suitable matching indicators in both data sources, the two files are linked by means of the Gower distance function which accounts for mixed binary and categorical measurement levels and is available in the `StatMatch` package in *R* (D'Orazio, 2009). Descriptive analyses do not indicate any occurrence of severe bias in the matched file.

[83] The precise alternatives were a recommendation for an intermediate track (Realschule) or the lower track (Hauptschule), but for statistical matching purposes, this variable is dichotomized.

[84] In contrast to *CHiSP*, *PHEO* did not test for students' intelligence test scores – which is why only grades were used for residualization purposes in both data sets regarding the performance dimension. Similarly, residuals are not net of students' self-assessment regarding their relative school performance due to lack of measurement in *PHEO*.

Using transition information from both sources, in the synthetic file, a dependent variable of subsequent educational transitions is computed measuring i) whether or not students made it to Gymnasium, ii) whether or not the remaining students passed Abitur on the first try, and iii) whether or not the remaining students started academic studies within 2 years. Comparing significance statistics for the prediction of transitions from primary to upper secondary school (Gymnasium) between original *PHEO* data on the one hand and the synthetic file on the other hand, results indicate potential overestimations of the effect of students' subjective expected probability of educational success, and also of the effect of subjective costs in the synthetic file. However, relatively similar values in both effect sizes and significance statistics can be observed for the three residuals that were used to identify self-fulfilling prophecies.

Comparing the impact of the residual terms on students' later educational transitions, I find significant but somewhat smaller effects on students' probability of passing Abitur – but insignificant estimates for students' propensity to start academic studies. Hence, based on the synthetic file, the general self-fulfilling prophecy hypothesis only holds for the first two transitions but has to be rejected for the third one.

Evaluating predicted probabilities for the comprehensive educational transition models obtained from sequential logit modeling, i) differences in predicted transition probabilities between students with primary school over- and underestimations, respectively, become smaller from transition 1 (+/- Gymnasium) to transition 2 (+/- Abitur); ii) differences between students with secondary school over- and underestimations, respectively, become smaller from transition 2 to transition 3 (+/- university transition); and iii) the predicted probabilities of students who have been overestimated at secondary school to make the transition to university are still higher than the predicted probabilities of students who had been overestimated at primary school. Observation i) and ii) give support for the hypothesis postulating decreasing teacher expectancy effects due to different treatments onto an immediately occurring educational transition over students' educational life course, and observation iii) gives support for the hypothesis postulating short-time teacher treatment effects to be larger than long-time effects.

Regarding robustness analyses, a comparison of the results obtained from the synthetic file with 'real' Panel data in terms of the British Cohort Study (*BCS*)[85] suggests that in the latter data, the same trend of differences in predicted probabilities can be observed. Additional sensitivity analyses (Buis, 2007, 2010, 2011) indicate that an unobserved but nonetheless confounding variable could be an issue of distortion particularly at later educational transition points. Notwithstanding this potential source of bias, I hold the view that in sum, results tend to support the assumption of a belief-mediated mechanism of Bayesian updating in a sequence of educational transitions.

[85] Although the BCS contains measures for both students' multiple educational transitions and multiple teachers' evaluations, its measures for student cost-benefit considerations are a far cry from being optimal. Hence, I still regard the results as obtained from statistical matching to be of distinct scientific value.

5 Conclusion

Summary Starting with Max Weber's definition of sociology as a scientific discipline that strives for both an *interpretive understanding*, and – by means of that – a *causal explanation* of its subject matter (Weber, 1964, p. 88), I first discussed how Weber attempts to unify the *Geisteswissenschaften* and positivism as two lines of thoughts that had separated from each other at the end of the 19^{th} century. In this context, I also highlighted that as regards the claim for a comprehensive understanding, Hempel's *deductive-nomological explanations* or *covering-law model* – that still prevails as the dominant paradigm in certain sociological textbooks (e.g. Opp, 2005b) – unfortunately falls behind both Weber's *understanding explanation* and Popper's situational logic entailing a relatively broad notion of rationality well-equipped to fulfill the demand of understanding social action.

Overcoming the methodological myopia of the covering-law model is also a crucial objective of recent *social-mechanism* approaches in analytical sociology. While it was shown that the existing multitude of definitions can be distinguished along the dimensions of i) their observability, ii) their law-likeness, and iii) their level of analysis, their general aims can best be described by Elster's metaphor of "opening the black box and showing the cogs and wheels of the internal machinery" (Elster, 1985, p. 5). In conjunction with a notion of subjective rationality of intermediate strength within strong bonds to situational restrictions (Goldthorpe, 1998) such as the *desires*, *beliefs*, and *opportunities* model proposed by Hedström (2005), *desire-mediated*, *belief-mediated* and *opportunity-mediated* mechanisms can be distinguished. Within that framework, following the maxim of *Occam's Razor* (Thorburn, 1918), a parsimonious theory of action should be preferred which may be amended by additional bridge assumptions where the simplified version is found to be unrealistic (Lindenberg, 1990). I postulated that this framework would also prove useful in approximating the causes and effects of teachers' evaluations by means of mechanism-based explanations in sense of a Weberian *understanding explanation*.

Having outlined these more abstract theoretical foundations, I then continued by linking Max Weber's (2000) definition of social class in terms of life chances to the thesis of *meritocracy* (Young, 1958; Bell, 1972) positing that individuals' social position should depend on their respective merits which are defined as the sum of individual IQ and effort. Having rejected the postulate of some authors (Jensen, 1969; Herrnstein and Murray, 1994) about the primacy of the heritability of intelligence and its impact on individual achievement compared to social background variables (Heckman, 1995; Fischer et al., 1996; Daniels et al., 1997; Korenman and Winship, 2000), I focused on both classic and recent theoretical approaches accounting for student *i*nequality in *e*ducational *o*pportunities (IEO). Grounding on a still influential distinction by Boudon (1974), social inequality in educational transitions can be decomposed into *primary effects* – mainly due to social-environmental factors – and *secondary effects* of social inequality – mainly due to varying utility considerations attached to a particular educational transition by parents of different social strata. To be precise, even if primary effects of social inequality in terms of performance differences between the social strata were accounted for, their educational transition propensities would still differ because of a *belief-mediated mechanism* of *relative risk aversion* (Breen and Goldthorpe, 1997): The higher parents' socioeconomic status, i) the higher their effort necessary to ensure their offspring would at least be equal off in terms of educational qualification; ii) the higher the expected benefit of higher education; iii) the higher the subjective expected probability of educational success (even apart from actual performance differences); and iv) the lower both monetary and transactional costs are perceived. This set the ground for the following analysis of teachers' evaluations as one of students' *significant others* affecting their life chances in terms of educational transition probabilities.

Finally, I tried to recapitulate the main findings of my four papers in this volume by using the theoretical toolbox discussed above. Focus of all four papers are teachers' evaluations of students' prospective academic aptitude as measured in the *Cologne High School Panel (CHiSP)*. While the first two papers consider these evaluations as an *outcome*, papers three and four use *residualized* teachers' evaluations – aimed to identify self-fulfilling prophecies in terms of a teacher treatment effect – as a predictor of the students' later educational transitions. Hence, in the 'applied' part of this introduction, I initially focused on the underlying mechanisms of the teacher side of the social situation as regards their formation of evaluations, and thereafter, on teachers' expectations – again measured in terms of their evaluations – as an intervening variable for students' cost-benefit considerations.

In the *first* paper, it is proposed that the formation of teachers' evaluations as measured in *CHiSP* can generally be reconstructed by frame-selection theory. The latter assumes that when individual actors are exposed to a social situation, at first, a certain *frame* of this situation will be activated either automatically, if unquestioned; or by means of more penetrative rational reflection, if problematic. While the frame of teachers' social situation – a survey question asking for a nonbinding evaluation of their students – will be unproblematic, and therefore, automatically activated, the *scripts* of action invoked in the second step might well vary between automatic processing and rational reflection. Most probably, teachers' beliefs about their students may be the result of an automatically-driven meritocratic assessment grounding on students' grades (*meritocracy-as-mode*). Perhaps less legitimate, but equally *understandable* might be teachers' automatic shaping of their evaluations based on student social background criteria, of which unconscious *habitus*-related match/mismatch considerations would be a prominent example. Complementary, both meritocratic criteria (such as students' intelligence apart from school grades) as well as social backgrounds (such as parental socioeconomic status as a perceived indicator for parents' ability to provide academic support) could also serve as *rationally-driven* motives of teachers' formation of beliefs about their students. Although empirical analyses based on *CHiSP* data are not suited to decompose automatically-driven and rationally-driven ways of teachers' belief formations, results indicate that meritocratic criteria such as student grades and intelligence are more important predictors of teachers' evaluations than parental socioeconomic status or student aspirations (as an indicator of what is usually referred to as an academic *habitus*) – though the latter variables are still significant. This could be interpreted in that teachers' evaluations are for the most part, though not perfectly, accurate beliefs of students' actual achievement.

The *second* paper asks to what extent teachers' evaluations would be subject to reference-group effects. Some arguments referred to what is conventionally known as *Big-Fish-Little-Pond Effect (BFLPE)* research (with students' academic self-concept as an outcome and class-level achievement as a contextual-level predictor), but since teachers' evaluations as an outcome are very distinct from the former dependent variable, also different social mechanisms had to be unfolded. While I argued that the conventional negative *BFLPE* of class-level achievement on students' academic self-concept is one *net of* a counterbalancing positive *Reflected-Glory Effect (RGE)* of school-level social status (Marsh et al., 2000) since two different social mechanisms are at work simultaneously (an *endogenous feedback effect* in case of the *BFLPE*, and an *exogenous contextual effect* in case of the *RGE*; see Manski, 1993, p. 31f.), a *negative* effect of class-level achievement on teachers' evaluations could be due to an *opportunity-mediated* mechanism in terms of teachers' adjustment of their reference standards independently of their beliefs about their students. Contrarily, either the *opportunity-mediated* mechanism of *regression to the mean* or the *belief-mediated mechanism* of *dissonance reduction* – conveniently also called *Halo* effect – could instead lead to a *positive* effect of class-level achievement on teachers' evaluations. In addition to these 'main effect' hypotheses, concatenations of opportunity- and belief-mediated mechanisms could account for interaction effects of class-level achievement with student-level achievement on the one hand, and with teachers' reported grading concepts on the other hand. Regarding results, multilevel analyses supported the hypotheses of a *Halo* effect caused by the concatenation of an opportunity- and belief-mediated mechanism of dissonance reduction – leading to a positive effect of class-mean achievement on teachers' evaluations. Concerning interaction effects, the coefficient of class-level aggregated student grade point average (GPA) significantly increases with better individual student marks.

Since results of the first two papers indicate that there is a residual part of inaccuracy in teachers' expectations as measured by teachers' evaluations, the *third* paper investigates to what extent this inaccuracy net of suitable measures of both student motivation and achievement could account for differences in students' educational transition propensities net of indicators of student (or parent) cost-benefit considerations regarding these transitions. The underlying mechanism of this hypothesized effect is one of student belief-mediation in terms of a *self-fulfilling prophecy* (Merton, 1948) due to a teacher treatment effect which has also been called the *Pygmalion* effect (Rosenthal and Jacobson, 1968). Theoretically, I assume this mechanism to operate via an unobserved effect of teacher 'caress' on students' later subjective expected probability of educational success that leads to differences in transition rates between students who have been overestimated and students who have been underestimated by their teacher. Empirically, this effect is identified by residualizing teachers' expectations (measured by their evaluations) on indicators of student achievement such as intelligence or grade point average as well as of motivation and academic self-concept. These residuals are found to significantly predict students' propensity to pass *Abitur* (German high school degree qualifying for academic studies) net of controls for expected benefit of education, the motive of status maintenance, subjective expected probability of educational success, and perceived costs of education. Furthermore, this effect is robust against two distinct approaches accounting for unobserved heterogeneity (Heckman, 1979; Buis, 2007, 2010, 2011). In contrast, no robust significant effect can be noted regarding students' university transitions – indicating that self-fulfilling prophecies in terms of a teacher treatment effect might diminish over time.

Exactly this is the crucial question of the *fourth* paper. In its theoretical section, first, several approaches from inequality in educational opportunity research suggesting decreasing effects of students' social backgrounds in their educational course of life – *Life Course Perspective* (Müller and Karle, 1993); *Maximally Maintained Inequality* (Raftery and Hout, 1993); *Relative Risk Aversion* (Breen and Goldthorpe, 1997); and *Effectively Maintained Inequality* (Lucas, 2001, 2009) – are discussed. Since on the one hand, *Relative Risk Aversion* is one of the two theories not rejected by Lucas (2009), and on the other hand, it explicitly links students' ability to their beliefs (in terms of subjective expected probabilities of educational success), it is considered to be the most promising candidate among the theories discussed in order to develop the argument for the hypothesis of decreasing *self-fulfilling prophecy* effects over time. Indeed, Breen and Goldthorpe (1999) and Breen and García-Peñalosa (2002) propose a belief-mediated mechanism of *Bayesian updating* that could well account for decreasing teacher treatment effects on students' beliefs about their expectations of success (as proposed for theoretical identification of self-fulfilling prophecy effects in paper 3). The difference in upward belief adjustment between students who have been overestimated and students who have been underestimated by their teachers – conditional on having successfully passed a transition, respectively – can be assumed to be positive. Therefore, I postulate i) that the effect of a teacher's expectation which is measured before the first transition point *on the first transition* is larger than the effect of a teacher's expectation which is measured before the second transition point *on the second transition*; and ii) that the long-term effect of a teacher's expectation measured before the first transition point *on the second transition* is smaller than the two possible short-term effects. Due to lack of suitable data, a synthetic file is created by means of *statistical matching* (D'Orazio et al., 2006) of the *CHiSP* one the one hand, and a panel survey named 'Elternhaus und Bildungschancen' (*Parental Home and Educational Opportunities; PHEO*) on the other hand – with *PHEO* comprising of measures of teachers' expectations, student achievement and motivation indicators, parent cost-benefit considerations in primary school; and *CHiSP* measuring respective variables in secondary school. Based on the resulting synthetic file observing artificial student transitions from primary to secondary school until starting academic studies, it is observed that i) differences in predicted transition probabilities between students with primary school over- and underestimations, respectively, become smaller from transition 1 (+/- Gymnasium) to transition 2 (+/- Abitur); ii) differences between students with secondary school over- and underestimations, respectively, become smaller from transition 2 to transition 3 (+/- university transition);

and iii) the predicted probabilities of students who had been overestimated at secondary school to make the transition to university are still higher than the predicted probabilities of students who had been overestimated at primary school. A set of robustness analyses indicates that these results are stable against a variety of statistical checks.

For illustrating purposes, figure 5 provides an overview of the crucial social mechanisms hypothesized for the four papers of the volume at hand. In the first paper, student characteristics A_s (such as achievement or social backgrounds) affect teachers' evaluations A_t passed on a *belief-mediated mechanism* B_t: For each of the constructs *meritocracy* and *habitus*, teachers' *variable rationality* Kroneberg (2005, 2006) is considered by assuming script selection either based on habitual routines and automatic processing (*as-mode*) or on rational reflection (*rc-mode).* At the current stage, the theoretical model remains silent on assumptions regarding potential variation in teachers' desires or their opportunity context.

This opportunity structure is explicitly considered in the second paper. Concretely, the classroom context O_t may directly 'lift' teachers' evaluations A_t in terms of reference-standard adjustments (*opportunity-mediated mechanism*), or also indirectly by means of an additional *belief-mediated mechanism* B_t when teachers project higher class mean achievement also on low-achievement students. Moreover, the classroom opportunity structure also moderates the belief-mediated mechanism of teachers' script selection: The higher school class' average achievement, counterintuitively, the more important achievement-related criteria become. Again, no assumptions are made regarding variation in teachers' desires.

In the third paper, the idea of a self-fulfilling prophecy in the educational system is reconstructed as a *concatenation* of *belief-mediated mechanisms.* First, student characteristics A_{s1} (such as social backgrounds, cf. Ready and Wright, 2011) invoke teachers' sometimes more positive, sometimes more negative beliefs B_t about their students.[86] According to the *Thomas theorem* (Thomas and Thomas, 1928), these beliefs may manifest in teacher treatment effects A_t'[87] such as classroom praise, bilateral encouragement, etc. The teacher treatment effect, in turn, has consequences for students' beliefs B_{s2} in terms of their subjective expected probability of educational success (p_{ep}). Well in line with the logic of Esser's (1999) SEU model, the self-fulfilling prophecy has consequences for students' educational transition propensities (here denoted as A_{s2}). As in paper one, I am agnostic about potential variation in both teachers' desires and opportunities.

The fourth paper entails only one change compared to the third one: Conditional on having passed a preceding transition A_{s2} or not, students upwardly or downwardly adjust their beliefs B_{s3} in terms of their subjected expected probability of educational success. According to Esser's (1999) SEU model, this affects students' subsequent transition propensities A_{s3}.

[86] In line with the paper's methodological specification, it is important to note that these are beliefs *apart from* both student achievement and motivation.

[87] I use the 'tick' in A_t' to clarify that the teacher treatment effect is analytically distinct from teachers' evaluations analyzed in papers 1 and 2 (though the same variable is used for empirical identification strategies).

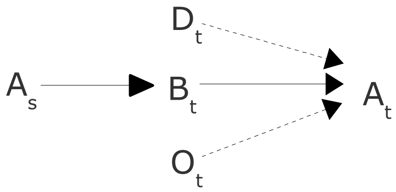
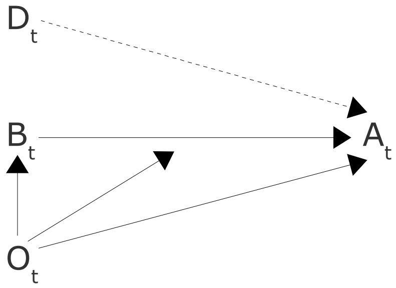
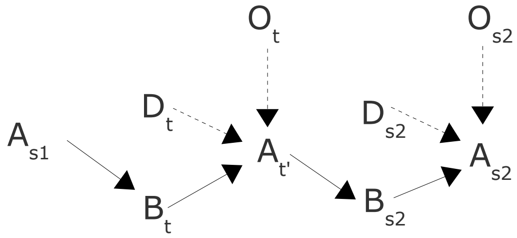
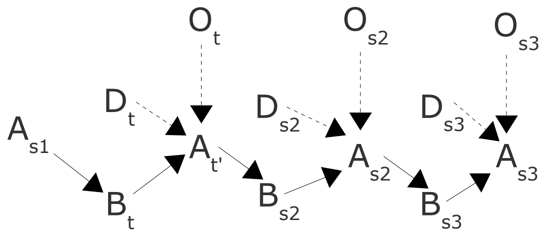
Paper	Social Mechanism(s)	Description
1) Teachers' Evaluations and the Definition of the Situation in the Classroom		A belief-mediated mechanism of script selection: Student backgrounds A_s invoke meritocracy- or background-related teacher scripts B_t in two modes of processing (automatic selection vs. rational choice).
2) Intelligence and Academic achievement as Contextual-Level Predictors of Teachers' Evaluations		A halo class-mean achievement effect on teachers' evaluations as a concatenation of an opportunity-mediated and a belief-mediated mechanism; plus an interaction effect in terms of teacher belief effects altered by varying opportunity structures.
3) The Impact of Teachers' Expectations on Students' Educational Opportunities in the Life Course		A self-fulfilling prophecy effect as a belief-mediated mechanism. Students' backgrounds A_{s1} invoke teachers' beliefs B_t, resulting in teacher treatment effects $A_{t'}$. These in turn affect student beliefs about their educational success B_{s2}, and thereby, their educational transitions A_{s2}.
4) Does the Effect of Teachers' Expectations on Students' Educational Opportunities Decrease over Educational Transitions?		An extended belief-mediated mechanism model of self-fulfilling prophecy effects. In addition to paper 3, it is assumed that a belief-mediated mechanism of Bayesian updating B_{s3} conditional of having successfully passed a preceding educational transition A_{s2} affects the next transition A_{s3}.

Figure 3: Summary of social mechanisms in the present volume.

Outlook Having set the ground for the papers included in this volume, finally, palpable limitations and resulting propositions for follow-up studies should be summarized. In **paper one**, a notable limitation is that due to data restrictions, we are not able to empirically identify the conditions of which mode of information-processing (automatic *as-mode* or rational *rc-mode*) is used by teachers. Hence, an urgent advice for future studies would be to develop even more fine-grained mechanism-based explanations in search of intervening opportunity structures or teacher belief sets that might account for differences in information processing. *Idealiter*, these new explanations should be tested by means of appropriate indicators leading to a better *understanding* of the information-processing side of teachers' expectancy formation. Methodologically, considering indicators on the teacher level would imply to extend our method of analysis towards a multilevel structural equation model (Muthén, 1994; Bauer, 2003) where teachers' evaluations of their students are nested in teacher contexts.

Another limitation of the first paper is that although a structural model is used as method of analysis, controls for measurement error may be regarded to be insufficient. This becomes particularly an issue concerning the operationalization of unconscious forms of symbolic communication between teachers and students usually referred to as effects of *habitus*: As indicated in the last section of this introduction, in the paper we found a significant effect of the analogy sub-dimension of intelligence on both students' average grade and teachers' evaluations. Within our framework distinguishing between *as-mode* and *rc-mode* processing on the one hand and meritocracy- and social-background-related predictors on the other hand, this could either be understood as a manifestation of a *meritocracy-rc-mode* type of processing (in terms of a particular importance of analogy-based inferences to the teachers). But on the other hand, since we do not dispose of a sufficiently high number of indicators in order to minimize measurement error, we cannot reject the alternative interpretation that teachers' appreciation of the verbal dimension of intelligence relates to *habitual* forms of student behavior that is valued by teachers in both their evaluations and grade assignments. Furthermore, on the one hand, the latter are also an aggregate form of teachers' evaluations (and could in turn be the result of *habitus*-related criteria); and on the other hand, students' *objective* academic performance could be understood as a latent variable which is only imperfectly measured by their average grades. For instance, teachers might refrain from considering students' class contribution all too much in their grading assignment (though one would expect a considerable correlation between the two), but class contribution might affect how teachers arrive at a more general evaluation of students' prospective aptitude for academic studies. Students' class contribution would now be a probable candidate to incorporate both *habitus*-related characteristics and to be responsible for the separate regression path of both students' average grades and teachers' evaluations on the verbal dimension of students' intelligence in terms of the analogy sub-score.

Regarding the latent variable of academic performance, a conventional first-order factor model will already be an improvement to current manifest approaches. Concerning measurement of a latent variable of *habitual* background-related forms of student behavior and symbolic communication, however, we would encourage future studies of teachers' evaluations to make use of a second-order factor model (Chen et al., 2005;

Rindskopf and Rose, 1988) wherein students' so-called *habitus* is the higher-order factor which is measured by lower-level factors such as parental SES, parental cultural capital (including cultural practices) and maybe also students' aspirations (but cf. footnote 63) – each of them ideally operationalized by a variety of appropriate indicators.

Regarding **paper two**, there is still work left concerning the theoretical justification of both 'main' and interaction effect hypotheses in the paper as well as with respect to their mechanism-based explanation in this introductory chapter. Thus, a more fine-grained analysis of the social mechanisms behind teacher-level variables such as their frames of reference as potential moderators of reference-group effects on teachers' evaluations will be inevitable. Following Hauser (1970a, p. 659), a thorough specification of the individual-level situation is of utmost importance before research can arrive at a comprehensive understanding also of contextual-level effects. Therefore, more research on the conditions of teachers' active mode of information-processing (*as-mode* vs. *rc-mode*) as demanded for the analysis of teachers' social situation in the classroom (paper 1) is also needed as a theoretical prerequisite for the analysis of contextual-level effects on teachers' evaluations and their interaction with teacher-level variables.

In this context, at least from my point of view, it is really astonishing that since Blalock and Wilken (1979), no effort has been invested in the analysis of opportunity-mediated mechanisms of individual cost-benefit considerations. From the more methodological literature (Friedrich, 1982; Brambor et al., 2006) we know that cross-level interaction effects can either be understood as as a contextual-level effect varying with an individual-level predictor, or as an individual-level effect varying with a contextual-level predictor – and the same, of course, applies for the social mechanisms behind. Hence, of course, the relative importance of student characteristics for a teacher to arrive at an evaluation may vary with class-average achievement – even if in our study, marginal effects of the complementary reading indicated relatively higher effects. Since an important limitation of our study is the selectivity of our data, we would not see the former reading to be rejected at all.

Even if the theoretical approach intentionally chosen here follows a very soft action-theoretical framework, future studies might wish to ground reference-group effect hypotheses – irrespectively of the outcome at hand – on more elaborate cost-benefit related social mechanisms. *Nota bene*, Blalock and Wilken (1979) already observed that "contextual variables may affect either subjective probabilities (expectancies) or utilities (values) and thereby produce either additive or multiplicative contextual terms" (p. 360). Since I do not see that sociologists have advanced in this respect in general, I recommend scholars to invest more theoretical effort in the specification of potentially opportunity-mediated mechanisms of social actors' utility considerations.

Furthermore, conventional *BFLPE* research found student gender composition in the classroom to be a moderator of class-level achievement effects on students' academic self-concept (Thijs et al., 2010). This effect can be explained by the assumption that individuals' beliefs about their abilities are shaped in relation to in-group standards (Rosenberg, 1979, ch. 6-9), and it was found that same-gender classmate beliefs more strongly affect students' academic decisions than opposite-gender classmate beliefs (Correll, 2001). For this reason, a higher negative *BFLPE* of same-gender classmate achievement than of opposite-gender classmate achievement on students' self-concept can be reconstructed as a concatenation of an opportunity- and a belief-mediated social mechanism. Similarly, teachers might hold different beliefs of their students depending on particular combinations of their own gender, students' gender, the gender composition in the classroom, and possibly even teachers' gender composition in school. Therefore, future studies should analyze if reference-group effects on teachers' evaluations vary with different gender contexts.[88]

Although a specification of the opportunity structure shaping students' utility considerations would certainly be an important issue for the theoretical model of self-fulfilling prophecies in terms of a teacher treatment effect, the more serious limitation of **paper three** can be seen in its *hiatus* between the *theoretical* and the *empirical* model. As outlined above, the identification assumption of the theoretical model is that self-fulfilling prophecies operate via both an indirect (in terms of student grades) and a direct unobserved teacher treatment effect on students' subjective expected probability of educational success – while all other cost-benefit parameters are assumed to be constant. I do not have to lay much emphasis on the fact that this assumption may be violated for a couple of reasons. Sensitivity analyses as performed in paper three try to address the objection that teachers dispose of private information on unobserved student characteristics that either enter their utility function or directly affect transition decisions – but that surely are only an imperfect approximation of potentially intervening variables. Hence, future studies should try to measure a larger set of indicators of teachers' evaluations in order to minimize measurement error and thereby the risk of neglecting teachers' private information about their students' academic aptitude. Empirically, this could be performed in well-known latent-variable frameworks such as confirmatory factor analysis and structural equation modeling, or also item response theory (Rasch, 1960; Embreston and Reise, 2000; Bond and Fox, 2001). A useful property of the latter approach – apart from separate estimation of person and item parameters – is that it also allows to test for the optimal treatment of empty categories such as students who never obtained a positive nor a negative evaluation by their teacher (i.e. whether they should be modeled in terms of an ordinal variable, or if they should rather be set to missing; cf. Drasgow and Hulin, 1990, p. 579).

Ideally, the theoretical model of self-fulfilling prophecies in terms of teacher-treatment effects operating via student differences in subjective expected probabilities of educational success should also be tested empirically. Hence, indicators for this crucial variable should be measured at different points in time to assess *changes* in subjective expected probabilities. If these either positive (in case of overestimations) or negative (in case of underestimations) changes were able to mediate the effect of the residual terms here used for empirical identification of self-fulfilling prophecies, a strong hint for the theoretical

[88] Another recommendation addressed in the paper but less interesting in the context of a theoretical résumé is that teachers' evaluations could be analyzed more comprehensively as a trivariate outcome by means of an ordered categorical multilevel model (Johnson, 1996, 1997; Gelman and Hill, 2007).

model's empirical validity would have been arrived at.

Moreover, as one limitation of conventional *Pygmalion* and self-fulfilling prophecy studies concerns an insufficient consideration of potential moderators (especially with regard to students' social backgrounds), future studies should also test for interaction effects of its respective measure of self-fulfilling prophecies with student (or parent) cost-benefit considerations. The latter, in turn, could be important intervening variables in terms of belief-mediated mechanisms that affect to what extent self-fulfilling prophecies actually take effect – resulting in another *concatenation* of social mechanisms.

And finally, although the theoretical framework used to explain self-fulfilling prophecies throughout this volume is one grounding on SEU theory, the social mechanisms that have been carved out for the first paper may have illustrated that rational comparisons of the alternatives at hand, as taken for granted in SEU theory, are only one very special case of actors' *variable rationality* (Kroneberg, 2006). To be precise, students' rather unconscious, automatic-spontaneous (*as-mode*) type of information processing should also be modeled in IEO research in general, and in self-fulfilling prophecy studies – that explicitly argue in favor of subtle and unconscious teacher treatment effects – in particular.

A mechanism-related limitation can also be identified in **paper four**. In its theoretical section, an explanation of *Bayesian* updating is proposed to account for decreasing self-fulfilling prophecy effects over students' educational transitions. However, these illustrations are, admittedly, a far cry from what will conventionally be understood as Bayesian modeling. Hence, to obtain a comprehensive belief-mediated account of decreasing self-fulfilling prophecy effects, social scientists should invest more effort in Bayesian models of belief updating. In my view, also sociologists' research about decreasing student background effects in their educational life course would benefit from this theoretical advancement.

But this limitation is certainly a more important issue regarding the consequences of *multiple* over- and underestimations for student belief updating. Madon et al. (2006) found that self-fulfilling prophecies can accumulate rather than dissipate over time if a whole series of over- and underestimations is considered. Unsurprisingly, Bayesian explanations of belief updating to explain accumulation of self-fulfilling prophecies still have to be developed.

In this context, also two *opportunity-mediated mechanism*s are worthwhile to be considered: First, it would be interesting to broaden the perspective on an analysis of whether the observed phenomenon of decreasing self-fulfilling prophecy effects over students' educational transition is in turn sensitive to temporal (i.e. period and cohort) or institutional effects. This is what *Maximally Maintained Inequality* (Raftery and Hout, 1993) postulated regarding the volatility of student background effects, and this could equally hold for self-fulfilling prophecy effects: Understanding the latter as teacher treatment effects on students' subjective expected probability of educational success, this concept could be a function of the variance in students' social composition, the student body's average ability in different grades *due to* differences in enrollment rates between student cohorts, periods or educational systems. Hence, future studies should pursue a more comparative perspective in order to be able to generalize the result of decreasing self-fulfilling prophecy effects.

Finally, another institutional (and thus opportunity-mediated) issue directly effects one measure of teachers' expectations used in paper four. In the primary-school sample (*Parental Home and Educational Opportunities; PHEO*) that is merged with the secondary-school sample (the *Cologne High School Panel; CHiSP*) via statistical matching (D'Orazio et al., 2006), primary school teachers' transition recommendations are used to operationalize teachers' expectations. However, German Federal states (*Bundesländer*) differ in the binding character of these recommendations as regards parental actual transitions: While at the moment, in Hessia, Mecklenburg-West Pomerania, North Rhine-Westphalia, Lower Saxony, Rhineland-Palatinate and the three German city states Hamburg, Berlin and Bremen, the final transition decision is left to the parents, in all other *Länder*, the educational system is permitted to correct parental decisions when differing from teachers' recommendations. Notably, these institutional differences have been observed to affect students' transition odds: A recent study by Gresch (2009) notes that differences in predicted transition probabilities to German highest-track secondary school (*Gymnasium*) between students of different social strata are lower in case of binding transition recommendations than in case of non-binding ones. At least two mechanism-based explanations could account for these differences: On the one hand, institutional constraints surely restrict parental *opportunities* to opt against a particular transition recommendation. On the other hand, institutional constraints could also alter teachers' *beliefs* about the extent to which their recommendation affects students' educational careers. In case of non-binding recommendations, teachers' and parents' beliefs about students' educational success show a partial interdependency in that teachers might try to anticipate how parents might decide in case of a particular recommendation. In case of binding recommendations, this anticipatory character is lessened since teachers know that parents may always be overruled. In consequence, differences in institutional constraints therefore affect the social selectivity of particular secondary school tracks in particular states. This might in turn influence students' subjective expected probability of educational success of the given secondary-school track; and hence, self-fulfilling prophecies could no longer be convincingly identified by means of that concept when, as in the case at hand, two samples from different Federal states with transition recommendations potentially differing in their binding character are used.[89]

To tell a long story short, future studies should first use an alternative measure of primary-school teachers' evaluations which is not affected by institutional constraints; and second, if future studies would still have to rely on evidence from statistically-matched files, ideally, two files from the same state should be used to exclude possible differences in students' cost-benefit considerations arising from differences in the social selectivity of a given state and thereby challenging theoretical identification of self-fulfilling prophecies as proposed in papers three and four.

[89]In order to compare the legal status of primary school teachers' transition recommendations from two different Federal states, access to their respective Education Act would be required – which has not been possible yet due to time constraints.

Concluding theoretical and methodological remarks In sum, research on teachers' evaluations (and, to a certain extent, also IEO research in general) would certainly benefit from a more thorough theoretical and methodological undergirding. Regarding theory, IEO research recently discovered the relevance of the framework of *counterfactuals* for the analysis of educational transitions (Erikson and Jonsson, 1996b; Erikson et al., 2005; Jackson et al., 2007). Furthermore, econometric theoretical models such as identification of a causal treatment effect have been adopted (for a review see Gangl, 2010). However, from my point of view, social scientists tend to use both concepts in a rather sloppy manner – completely ignoring the ontological and epistemological prerequisites involved. I do not have to lay much emphasis on the fact that the notion of *causality* is one of the philosophical problems discussed most intensely since ancient Greek philosophy (see e.g. Gotthelf, 1976). For statisticians such as Holland (1986) or Rubin (1986), manipulability theories of causation are particularly attractive for their general assumption that manipulation of a cause will result in the manipulation of an effect (Woodward, 2008). But even these largely pragmatic theories of causation are not without problems: For instance, von Wright's (1971) account relates p's bringing about q to an agency concept that would have difficulties to incorporate environmental effects – that might also affect human interaction.[90] And even the more recent and elaborated structural theory by Judea Pearl (2000) is not immune against the objection of circularity (Woodward, 2008).[91] Hence, more effort should be invested into the specification of the underlying causal framework of the social sciences.

Similarly, *counterfactuals* are much more than a methodological tool-box easily at hand for educational transition analysts. As Menzies (2009) points out, a notion of causality that would be interpreted as counterfactual from today's point of view was already held by David Hume[92], and as indicated above, this notion culminated in David Lewis' very elaborated theory of counterfactual causality (Lewis, 1973, 1977, 1979, 1981). However, it is important to note that Lewis' condition "If it were the case that A, then it would be the case that C" (Lewis, 1973, p. 418) is more and more relaxed throughout his paper to be true also when some A-world where C holds is closer to the actual world than is any A-world where C does not hold (Menzies, 2009). This means that Lewis holds a realism also about *non-actual possible* worlds close to ours that does not necessarily coincide with social scientists' pragmatic usage of methodological counterfactualism as a tool for decomposing primary and secondary effects of social inequality.

I do not say that methodological counterfactualism – given its ontological foundation – is not suited for mechanism-based explanations that also strive to ensure a deeper

[90] An illustrative example for an environmental effect considerably constraining individual interaction is the war in Dafur: A progressive desertification and soil erosion for the last 30 years caused a decrease in both cultivable areas and grasslands. This made members of several ethnic groups to leave their ancestral regions and to move to the higher-precipitation areas in the south of Sudan – where they came into conflict with local ethnic groups (University for Peace, 2004).

[91] For an application of Pearl's theory of causation on educational inequality see Morgan and Winship (2012).

[92] "We may define a cause to be an object followed by another [...] where, if the first object had not been, the second never had existed" (Hume, 1913, section VII).

understanding.[93] Quite contrarily, the founding father of a unified approach in this respect, Max Weber, holds a notion of causation that is notably close to contemporary counterfactualism:

> "The judgement that if a single historical fact is conceived of as absent from or modified in a complex of historical conditions, it *would* condition a course of historical events in such a way which would be different in certain *historically important* respects, seems to be of considerable value for the determination of the 'historical significance' of those facts" (Weber, 1948, p. 166; orig. emph.).[94]

Thus, in sum, social scientists should avoid overhasty usage of concepts fraught with meaning that is not entirely grasped, and they should devote more work in the theoretical underpinnings of those concepts that are considered to be useful from a pragmatic point of view.

Another issue quite relevant for rational-choice or subjective-expected-utility-based explanations of social inequality in educational opportunities in general affects the underlying decision heuristics. As outlined above, current theoretical accounts assume that the underlying decision heuristic reads "Maximize the expected value!" equally for all social strata (Esser, 1999, p. 274). However, social psychologists have revealed that individuals not always act according to this admittedly efficient strategy – but may in particular situations have good reasons to follow alternative heuristics that are mediocre from a strict point of view (Simon, 1955, 1957; Tversky and Kahneman, 1974; also see Coombs et al., 1970, ch. 5; Lave and March, 1975, pp. 140-143). For instance, inequalities in educational opportunities would probably increase if parents from the lower strata would act according to the comparably conservative *maximin* principle – maximizing the minimum possible value while completely ignoring subjective probabilities even if they may be relatively high – while middle- and higher-class parents still maximize their expected value.[95] Hence, IEO theory should, possibly by means of agent-based modeling (Hedström, 2005, ch. 6), evaluate the consequences of different constellations of class-variant decision heuristics when all other parameters are held constant. This, of course, would amount to soften Occam's razor in introducing additional auxiliary assumptions.

Regarding methods, in particular realist generative process theorists as referred to in the theoretical section would admittedly insist (or at least recommend) that causal forces could best be approximated by means of small or medium N studies (Gerring, 2007). This advice intuitively makes sense since the more fine-grained the mechanism-based explanation (i.e. the smaller the entities the observed factorial regularity is decomposed into), the harder it is to get closer to the actual causal structure by observable

[93] Once the above-noted issues have been resolved, counterfactual thinking as a *decision heuristic* might prove helpful in both clarifying the efficacy of contrast effects in reference-group effect research (Roese, 1994) and mediating individuals' expectancies of success through subjective perceptions of control (Roese, 1994; Nasco and Marsh, 1999).

[94] See Ringer (2002) for a more elaborate discussion of Weber's notion of causation.

[95] See Jæ ger and Holm (2012) for a first attempt of distinguishing between 'conformists' and 'rebels' in the framework of Breen and Goldthorpe's (1997) relative risk aversion theory.

quantitative indicators. Hence, the set of social mechanisms providing an *understanding explanation* of the causes and effects of teachers' evaluations as reconstructed in this introduction should *idealiter* be evaluated by means of *process tracing* in small or medium N studies (Mahoney, 2000; Bennett and Elman, 2006; Bennett, 2010; Collier, 2011; Beach and Pedersen, 2012; Rohlfing, 2012, ch. 6).

Sticking to the quantitative framework as applied here, recent evidence would suggest that listwise deletion of missing values as applied in the four papers at hand is usually inferior in precision compared to methods of *imputation* (Schafer, 1997; Schafer and Graham, 2002; Graham, 2009). To be precise, statisticians distinguish between three scenarios of missing data: missing at random (*MAR*), missing completely at random (*MCAR*), and missing not at random (*MNAR*). *MAR* allows the probabilities of missingness to depend on the observed data but not on the missing data. *MCAR* is a special case of *MAR* where the probability of missingness does not depend on the observed data either. Consequently, *MNAR* already occurs when the *MAR* condition is violated.

Unless designs of planned missingness such as cohort-sequential longitudinal designs or administration of multiple questionnaires with varying subsets of items occur, violation of *MCAR* or *MAR* can be expected to be likely (Schafer, 1997, ch. 2); and results of statistical analyses obtained after listwise deletion may be biased if the data is not *MCAR* or at least *MAR* (Schafer and Graham, 2002).

In contrast, Graham and Donaldson (1993) as well as Collins et al. (2001) found that in many realistic scenarios violating *MAR*, the bias potentially resulting by falsely assuming *MAR* will actually be negligible. Graham (2009, p. 554) adds that especially in case of multiple regression models, the most undesirable property of listwise deletion is its loss of power. To be sure, listwise deletion is superior to pairwise deletion since as in contrast to the latter, a common set of cases for all analyses is used (cf. Schafer, 1997, p. 39).[96] Single imputation methods such as arithmetic mean imputation (Wilks, 1932), conditional mean imputation (Buck, 1960) or stochastic regression imputation – augmenting each score predicted by conditional mean imputation with a normally distributed residual term (cf. Enders, 2010, ch. 2.8.) – are known to underestimate standard error in regression analyses (Enders, 2010, p. 48).

Full information maximum likelihood (*FIML*) that uses the whole set of available cases in order to obtain ML estimates of the aspired values is known to behave well under *MAR*, but not under *MNAR* conditions (Enders, 2010, p. 87). As one crucial point of arguing of papers three and four is to tackle the objection of potential unobserved heterogeneity, one might arrive at the conclusion that there is no well-thought reason to reject the assumption of *MNAR a priori*. Furthermore, paper one used an input matrix of polychoric correlations (Olsson, 1979; Muthén, 1984; Aish and Jöreskog, 1990; Jöreskog, 1994) in order to account for categorical outcomes in the structural model. In contrast, using the *FIML* approach to impute missing values would require the use of raw data (Enders, 2010, p. 123).

[96] The problem of inconsistent sets of cases is particularly an issue for statistical approaches relying on a variance-covariance matrix as data input such as structural equation modeling (Little, 1992; Marsh, 1998; Wothke, 1993) – which is the method of analysis of paper 1 in the volume at hand.

The only imputation technique that is approximately stable under *MNAR* conditions is *multiple imputation (MI)*. Basically, *MI* starts from stochastic regression imputation but then generates multiple copies of the underlying data set, each filled with different missing values. The analyses of interest are then performed simultaneously with all different data, and these results are pooled – typically by means of the data augmentation algorithm (Schafer, 1997; Tanner and Wong, 1987) – in order to arrive at one final estimate. The stochastic regression imputation step predicts the values for a variable to be imputed by regressing it on various predictors and then adds a normalized residual term obtained from this regression equation.[97] Now recall that papers three and four use the residuals estimated from a logistic regression of teachers' expectations on students' educational performance and motivation to arrive at an empirical identification of self-fulfilling prophecies. In my view, it has to be evaluated first whether the stochastic regression part of multiple imputation does neither upwardly nor downwardly bias estimation strategies that also make use of residualized variables.[98] Furthermore, there is no option to estimate the polychoric correlation matrix that has been used as input matrix for the confirmatory factor analysis and structural equation models as performed in paper 1 in an MI framework.[99] Finally, I see the risk of generally augmenting bias in the data when missing variables are first imputed and then used for data fusion via statistical matching. Therefore, I stick to the conventional approach of listwise deletion since it is unsettled whether the more sophisticated techniques would not rather lower the precision of the estimates.

Apart from potential concerns with missing data handling that I see rejected by the above-specified arguments, a general objection against particular methodological strategies such as statistical matching (D'Orazio et al., 2006) could be their peril. However, on the one hand, from an epistemological point of view, I follow Popper (1963) in his opinion that researchers should always formulate hypotheses as falsifiable as possible in order to enable empirical research to reduce complexity in that only those hypotheses unrejected so far are corroborated for the time being.[100]

On the other hand, from a more pragmatic point of view, when the alternatives comprise doing something risky (maybe even transcending the Popperian sense of it),

[97] I above said that MI is only *approximately* stable under *MNAR* conditions since the imputation process has to be corrected as proposed by Rubin (1987): "Generate multiple imputations under an MAR mechanism, then add a constant to the imputed values to compensate for the possibility that the MAR-based imputations may be too high or too low" (Enders, 2010, p. 289).

[98] A more practical reason against the feasibility of multiple imputation throughout paper 3 is that the Inverse Mill's Ratios for each SEU predictor and residual term to control for heterogeneity bias had to be computed by hand. Hence, it wouldn't have been possible to estimate a Heckman model simultaneously on several imputed data sets. In fact, the selection model applied in paper 3 can be regarded as a (admittedly imperfect) method to account for missingness under *MNAR* conditions (cf. Enders, 2010, pp. 291-298).

[99] Asparouhov and Muthén (2010) showed that given a sufficiently high number of imputed data sets (5 and more), *Weighted Least Squares* may yield consistent results in both single-level and two-level structural equation models with categorical variables.

[100] "A serious empirical test always consists in the attempt to find a refutation, a counter example (sic!). In the search for a counter example, we have to use our background knowledge; for we always try to refute first the *most risky* predictions (...)" (Popper, 1963, p. 240; orig. emph.).

or discarding analyses completely, I take the liberty to opt for the former. As Boudon (1976, p. 1185) prominently remarks: "But I always thought that 'try, and see what happens' was a better though less secure line of behavior as far as research was concerned than following a well-explored trail – for instance, applying to data statistical techniques which are mechanically well-known and which always 'work'." Pursuant to the title of an essay by Giuliani (2003), the attempt of the volume at hand is to follow *a sociological methodology that does not attempt to understand more than it can explain, and that does only explain what has been understood* – with an associated demand of being on the safe side theoretically wherever daring methodologically.

References

Adorno TW; Albert H; Dahrendorf R et al. (1976) *The Positivist dispute in German sociology.* New York / Hagenstown, San Francisco, London: Harper and Row.

Aigner DJ (1970) A comment on problems in making inferences from the Coleman Report. *American Sociological Review* 35(2): 249–252.

Aish AM and Jöreskog KG (1990) A panel model for political efficacy and responsiveness: an application of LISREL 7 with weighted least squares. *Quality and Quantity* 24(4): 405–426.

Albert H (1985) *Treatise on Critical Reason.* Princeton, New Jersey: Princeton University Press.

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

Apel KO (1979) *Die Erklären:Verstehen-Kontroverse in transzendentalpragmatischer Sicht.* Frankfurt a.M.: Suhrkamp.

Asparouhov T and Muthén BO (2010) Weighted least squares estimation with missing data. *Mplus Technical Appendix* .
URL: *http://www.statmodel.com/download/GstrucMissingRevision.pdf*

Ayer AJ (1952) *Language, Truth, and Logic.* New York: Dover Publications.

Balog A (2008) Verstehen und Erklären bei Max Weber. In: Greshof R; Kneer G and Schneider WL (eds.) *Verstehen und Erklären - Sozial- und kulturwissenschaftliche Perspektiven*, 73–93. München: Wilhelm Fink Verlag.

Bandura A (1969) Social-learning theory of identificatory processes. *Handbook of socialization theory and research* 213–262.

Barton AH (1970) Comments on Hauser's "Context and Consex". *American Journal of Sociology* 76(3): 514–517.

Bauer DJ (2003) Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics* 28(2): 135–167.

Beach D and Pedersen RB (2012) Process-Tracing Methods .

Bechtel W (2006) *Discovering Cell Mechanisms. The Creation of Modern Cell Biology.* New York: Cambridge Univ Press.

Bechtel W (2008) *Mental Mechanism. Philosophical Perspectives on Cognitive Neuroscience.* London: Routledge.

Bechtel W and Abrahamsen A (2005) Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36(2): 421–441.

Beck PA (1991) Voters' intermediation environments in the 1988 presidential contest. *Public Opinion Quarterly* 55(3): 371–394.

Beck U (1986) *Risikogesellschaft. Auf dem Weg in eine andere Moderne.* Frankfurt a.M.: Suhrkamp.

Beck U (1992) *Risk Society. Towards a New Modernity.* London/Thousand Oaks/New Dheli: Sage Publications.

Becker D and Birkelbach K (2010) Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond- oder Reflected-Glory-Effekt? Eine Mehrebenen-Analyse von Lehrerurteilen. In: Beckers T; Birkelbach K; Hagenah J et al. (eds.) *Komparative empirische Sozialforschung*, 113–141. Wiesbaden: VS-Verlag.

Becker G (1976) *The Economic Approach to Human Behavior.* Chicago: Chicago University Press.

Becker GS (1962) Investment in human capital: a theoretical analysis. *The Journal of Political Economy* 70(5): 9–49.

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 (2012) Bildungsungleichheit im Lichte aktueller Theorieanwendung in der soziologischen Bildungsforschung. *Bildung - Arbeit - Erwachsenwerden. Ein interdisziplinärer Blick auf die Transition im Jugend und jungen Erwachsenenalter* 43–75.

Becker R and Hadjar A (2011) Meritokratie - Zur gesellschaftlichen Legitimation ungleicher Bildungs-, Erwerbs-und Einkommenschancen in modernen Gesellschaften. *Lehrbuch der Bildungssoziologie* 37–62.

Becker R and Hecken AE (2009a) 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.

Becker R and Hecken AE (2009b) 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 Schubert F (2006) Soziale Ungleichheit von Lesekompetenzen. Eine Matching-Analyse im Längsschnitt mit Querschnittsdaten von PIRLS 2001 und PISA 2000. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58(2): 253–284.

Bell D (1972) On equality: I. Meritocracy and equality. *Public Interest* 29: 29–68.

Benjamin W (1968) Theses on the Philosophy of History. In: Arendt H (ed.) *Walter Benjamin: Illuminations*, 253–264. New York: Schocken Books.

Bennett A (2010) Process tracing: A Bayesian perspective. *The Oxford handbook of political methodology* 702–721.

Bennett A and Elman C (2006) Complex causal relations and case study methods: the example of path dependence. *Political Analysis* 14(3): 250–267.

Bernstein B (1971) *Class, Codes and Control. Vol I: Theoretical Studies towards a Sociology of Language*. London: Routledge.

Bernstein B (1981) Codes, modalities, and the process of cultural reproduction: A model. *Language in society* 10(3): 327–363.

Bernstein B (1990) *Class, Codes and Control. Vol. IV: The Structuring of Pedagogic Discourse*. London: Routledge.

Bernstein B (1995) Code theory and its positioning: a case study in misrecognition. *British Journal of Sociology of Education* 16(1): 3–19.

Bhaskar R (1975) *A Realist Theory of Science*. Leeds: Leeds Books.

Biggs M (2009) Self-fulfilling prophecies. In: Baerman P and Hedström P (eds.) *The Oxford Handbook of Analytical Sociology*, 294–314. Oxford: Oxford University Press.

Birkelbach K (2011) Teacher evaluations over the life course: valid prognosis or self-fulfilling prophecy? In: Mica A; Peisert A and Winczorek J (eds.) *Sociology and the Unintended - Robert Merton Revisited*, 133–153. Frankfurt a.M.: Peter Lang.

Blalock HM (1984) Contextual-effects models: theoretical and methodological issues. *Annual Review of Sociology* 10: 353–372.

Blalock HM and Wilken PH (1979) *Intergroup Processes. A Micro-Macro Perspective.* New York: The Free Press.

Blau PM and Duncan OD (1967) *The American Occupational Structure.* New York: John Wiley.

Blossfeld HP and Shavit Y (1993) Persisting barriers. In: Blossfeld HP and Shavit Y (eds.) *Persistent Inequality - Changing Educational Attainment in Thirteen Countries*, 1–23. Boulder, CO: Westview Press.

Böhm JM (2008) Verstehen und Erklären bei Popper. In: Greshoff R; Kneer G and Schneider L (eds.) *Verstehen und Erklären - Sozial- und kulturwissenschaftliche Perspektiven*, 365–389. München: Wilhelm Fink Verlag.

Bond TG and Fox CM (2001) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bornmann L (2010) Die analytische Soziologie: Soziale Mechanismen, DBO-Theorie und Agentenbasierte Modelle. *Österreichische Zeitschrift für Soziologie* 35(4): 25–44.

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

Boudon R (1976) Comment on Hauser's review of Education, Opportunity, and Social Inequality. *The American Journal of Sociology* 81(5): 1175–1187.

Boudon R (1979) Generating models as a research strategy. In: Merton RK; Coleman JS and Rossi PH (eds.) *Qualitative and Quantitative Social Research*, 51–64. New York: The Free Press.

Boudon R (1981) *The logic of social action: an introduction into sociological analysis.* London: Routledge & Kegan Paul.

Boudon R (1991) What middle-range theories are. *Contemporary Sociology: An International Journal of Reviews* 20(4): 519–522.

Boudon R (1996) The 'cognitivist model': a generalized rational-choice model. *Rationality and society* 8(2): 123–150.

Boudon R (1998) Limitations of rational choice theory. *American Journal of Sociology* 104(3): 817–828.

Boudon R (2003) Beyond rational choice theory. *Annual review of sociology* 29: 1–21.

Bourdieu P (1973) Cultural reproduction and social production. In: Brown R (ed.) *Knowledge, Education, and Cultural Change: Papers in the Sociology of Education*, 71–112. London: Tavistock Publications.

Bourdieu P (1977) *Outline of a Theory of Practice (Cambridge Studies in Social and Cultural Anthropology).* Cambridge: Cambridge University Press.

Bourdieu P (1986) *Distinction: A Social Critique of the Judgement of Taste.* London: Routledge.

Bourdieu P (1988) Vive la crise! *Theory and Society* 17(5): 773–787.

Bourdieu P (1991) *Language and Symbolic Power.* Cambridge: Polity Press.

Bourdieu P and Passeron JC (1990) *Reproduction in Education, Society and Culture.* London: Sage Publications Ltd.

Bowles S and Levin HM (1968) The determinants of scholastic achievement - an appraisal of some recent evidence. *Journal of Human Resources* 3(1): 3–24.

Brambor T; Clark WR and Golder M (2006) Understanding interaction models: improving empirical analyses. *Political Analysis* 14(1): 63–82.

Braun N (1995) Individual thresholds and social diffusion. *Rationality and Society* 7(2): 167–182.

Breen R (1996) *Regression Models: Censored, Sample Selected or Truncated Data.* Thousand Oaks, CA: Sage Publications.

Breen R (1999) Beliefs, rational choice and bayesian learning. *Rationality and Society* 11(4): 463–479.

Breen R and García-Peñalosa C (2002) Bayesian learning and gender segregation. *Journal of Labor Economics* 20(4): 899–922.

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 Goldthorpe JH (1999) Class inequality and meritocracy: a critique of Saunders and an alternative analysis. *The British Journal of Sociology* 50(1): 1–27.

Breen R and Jonsson JO (2000) Analyzing educational careers: a multinomial transition model. *American Sociological Review* 65(5): 754–772.

Breen R; Luijkx R; Müller W et al. (2009) Nonpersistent inequality in educational attainment: evidence from eight European countries. *American Journal of Sociology* 114(5): 1475–1521.

Breen R and Rottman D (1995) Class analysis and class theory. *Sociology* 29(3): 453–473.

Briggs DC (2004) Causal inference and the Heckman model. *Journal of Educational and Behavioral Statistics* 29(4): 397–420.

Bryk AS and Raudenbush SW (1992) *Hierarchical Linear Models.* Thousand Oaks, CA: Sage Publications.

Buck SF (1960) A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society. Series B (Methodological)* 22(2): 302–306.

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.

Buis ML (2011) The consequences of unobserved heterogeneity in a sequential logit model. *Research in Social Stratification and Mobility* 29(3): 247–262.

Bunge M (1997) Mechanism and explanation. *Philosophy of the Social Sciences* 27(4): 410–465.

Bunge M (2004) How does it work? *Philosophy of the Social Sciences* 34(2): 182–210.

Byrne BM; Shavelson RJ and Muthén BO (1989) Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin* 105(3): 456–466.

Cain GG and Watts HW (1968) The controversy about the Coleman report: comment. *The Journal of Human Resources* 3(3): 389–392.

Cain GG and Watts HW (1970) Problems in making policy inferences from the Coleman Report. *American Sociological Review* 35(2): 228–242.

Carver RP (1975) The Coleman Report: using inappropriately designed achievement tests. *American Educational Research Journal* 12(1): 77–86.

Chen FF; Sousa KH and West SG (2005) Teacher's corner: testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal* 12(3): 471–492.

Cialdini RB; Borden RJ; Thorne A et al. (1976) Basking in reflected glory: three (football) field studies. *Journal of Personality and Social Psychology* 34(3): 366–375.

Coleman JS (1966) *Equality of Educational Opportunity.* Washington, DC: U.S. Government Printing Office.

Coleman JS (1968) Equality of Educational Opportunity: reply to Bowles and Levin. *The journal of human resources* 3(2): 237–246.

Coleman JS (1970) Reply to Cain and Watts. *American Sociological Review* 35(2): 242–249.

Coleman JS (1990) *Foundations of Social Theory.* New York: Harvard University Press.

Collier A (1989) *Scientific Realism and Socialist Thought.* Hemel Hempstead: Harvester Verso.

Collier A (1994) *Critical Realism: An Introduction to Roy Bhaskar's Philosophy.* London: Verso.

Collier D (2011) Understanding process tracing. *PS: Political Science & Politics* 44(4): 823–830.

Collins LM; Schafer JL and Kam CM (2001) A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 6(4): 330–351.

Collins RL (1996) For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin* 119(1): 51–69.

Comte A (1839) *Cour de philosophie positive*, vol. 4. Paris: Bachelier.

Comte A (1865) *A General View Of Positivism.* London: Trübner and Co.

Coombs CH; Dawes RM and Tversky A (1970) *Mathematical Psychology: An Elementary Introduction.* Englewood Cliffs, NJ: Prentice Hall.

Correll SJ (2001) Gender and the career choice process: the role of biased self-assessments. *American Journal of Sociology* 106(6): 1691–1730.

Craver CF (2007) *Explaining the Brain. Mechanisms and the Mosaic Unity of Neuroscience.* Oxford: Oxford University Press.

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.

Dai DY and Rinn AN (2008) The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review* 20(3): 283–317.

Daniels M; Devlin B and Roeder K (1997) Of genes and IQ. In: Devlin B; Fienberg SE; Resnick DP et al. (eds.) *Intelligence, Genes, & Success: Scientists Respond to the Bell Curve*, 45–70. New York: Springer.

Daniels N (1978) Merit and meritocracy. *Philosophy & Public Affairs* 7(3): 206–223.

Darden L (2006) *Reasoning in Biological Discoveries. Essays on Mechanisms, Interfield Relations, and Anomaly Resolution.* Cambridge, UK: Cambridge University Press.

Davidson D (1980) *Essays on Actions and Events.* Oxford: Clarendon Press.

Davis JA (1966) The campus as a frog pond: an application of the theory of relative deprivation to career decisions of college men. *The American Journal of Sociology* 72(1): 17–31.

Davis K and Moore WE (1945) Some principles of stratification. *American Sociological Review* 10(2): 242–249.

Dijkstra P; Kuyper H; van der Werf G et al. (2008) Social comparison in the classroom: a review. *Review of Educational Research* 78(4): 828–879.

Dilthey W (1883) *Einleitung in die Geisteswissenschaften: Versuch einer Grundlegung für das Studium der Gesellschaft und der Geschichte.* Leipzig: Duncker & Humblot.

Dilthey W (1924) Ideen über eine beschreibende und zergliederte Psychologie. In: Dilthey W (ed.) *Die geistige Welt*, 139–240. Leipzig und Berlin: Teubner.

Dilthey W (1927) *Der Aufbau der geschichtlichen Welt in den Geisteswissenschaften.* Leipzig und Berlin: Teubner.

Dilthey W (1984) Introduction to the Human Sciences. In: Makkreel RA and Rodi F (eds.) *Selected Works (Vol. I).* Princeton, New Jersey: Princeton University Press.

DiMaggio P (1982) Cultural capital and school success: the impact of status culture participation on the grades of U.S. high school students. *American Sociological Review* 47(2): 189–201.

DiPrete TA and Eirich GM (2006) Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology* 32: 271–297.

D'Orazio M (2009) *StatMatch: Statistical Matching. R package version 0.8.*
URL: *http://cran.r-project.org/package=StatMatch*

D'Orazio M; Di Zio M and Scanu M (2006) *Statistical Matching - Theory and Practice.* Chichester: John Wiley & Sons.

Drasgow F and Hulin CL (1990) Item response theory. In: Dunette MD and Hough LM (eds.) *Handbook of industrial and organizational psychology*, vol. 1, 2nd edn., 577–636. Consulting Psychologists Press Palo Alto, CA.

Droysen JG (1857) Grundriss der Historik. In: Droysen JG (ed.) *Historik (1977, ed. by Peter Leyh).* Stuttgart, Bad Cannstatt: Frommann-Holzboog.

Dumais SA (2006) Early childhood cultural capital, parental habitus, and teachers' perceptions. *Poetics* 34(2): 83–107.

Elster J (1983) *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press.

Elster J (1985) *Making Sense of Marx.* Cambridge: Cambridge University Press.

Elster J (1989) *Nuts and Bolts for the Social Sciences.* Cambridge: Cambridge University Press.

Elster J (1998) A plea for mechanisms. In: Hedström P and Swedberg R (eds.) *Social mechanisms: An analytical approach to social theory*, 45–73. Cambridge: Cambridge University Press.

Elster J (1999) *Alchemies of the Mind: Rationality and the Emotions.* Cambridge: Cambridge University Press.

Elster J (2007) *Explaining Social Behavior - More Nuts and Bolts for the Social Sciences.* Cambridge: Cambridge University Press.

Embreston SE and Reise SP (2000) *Item Response Theory for Psychologists.* Mahwah, NJ: Lawrence Erlbaum.

Enders CK (2010) *Applied Missing Data Analysis.* New York: The Guilford Press.

Erbring L and Young AA (1979) Individuals and social structure: contextual effects as endogenous feedback. *Sociological Methods & Research* 7(4): 396–430.

Erikson RS; Goldthorpe JH; Jackson M et al. (2005) On class differentials in educational attainment. *Proceedings of the National Academy of Sciences of the United States of America* 102(27): 9730–9733.

Erikson RS and Jonsson JO (1996a) Introduction: explaining class inequality in education: the Swedish test case. In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 1–63. Boulder, CO: Westview Press.

Erikson RS and Jonsson JO (1996b) The Swedish context: educational reform and long-term change in educational inequality. In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 65–93. Boulder, CO: Westview Press.

Erikson RS and Rudolphi F (2010) Change in social selection to upper secondary school - primary and secondary effects in Sweden. *European Sociological Review* 26(3): 291–305.

Esser H (1993a) *Soziologie. Allgemeine Grundlagen.* Frankfurt / New York: Campus Verlag.

Esser H (1993b) The Rationality of Everyday Behavior: A Rational Choice Reconstruction of the Theory of Action by Alfred Schutz. *Rationality and Society* 5(1): 7–31.

Esser H (1996a) Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48(1): 1–34.

Esser H (1996b) What is wrong with 'variable sociology'? *European Sociological Review* 12(2): 159–166.

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

Esser H (2010) Das Modell der Frame-Selektion. Eine allgemeine Theorie der Sozialwissenschaften. In: Albert G and Sigmund S (eds.) *Soziologische Theorie kontrovers (Sonderheft 50 der Kölner Zeitschrift für Soziologie und Sozialpsychologie)*, 50–62. Wiesbaden: VS Verlag für Sozialwissenschaften.

Esser H and Kroneberg C (2010) Am besten nichts Neues? In: Albert G and Sigmund S (eds.) *Soziologische Theorie kontrovers (Sonderheft 50 der Kölner Zeitschrift für Soziologie und Sozialpsychologie)*, 79–86. Wiesbaden: VS Verlag für Sozialwissenschaften.

Eysenck HJ (1975) Equality and education: fact and fiction. *Oxford Review of Education* 1(1): 51–58.

Farkas G (1974) Specification, residuals and contextual effects. *Sociological Methods & Research* 2(3): 333–363.

Feldman L and Price V (2008) Confusion or enlightenment? How exposure to disagreement moderates the effects of political discussion and media use on candidate knowledge. *Communication Research* 35(1): 61–87.

Festinger L (1957) *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Feuer LS (1957) The principle of simplicity. *Philosophy of Science* 24(2): 109–122.

Fischer CS; Hout M; Jankowski MS et al. (1996) *Inequality by design.* Princeton, New Jersey: Princeton University Press.

Freedman DA (1987) As others see us: a case study in path analysis. *Journal of Educational and Behavioral Statistics* 12(2): 101–128.

Friedrich RJ (1982) In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science* 26(4): 797–833.

Galton F (1869) *Heredity Genius: An Inquiry Into Its Laws and Consequences*. London: Collins.

Galton F (1886) Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15: 246–263.

Gambetta D (1998) Concatenations of mechanisms. In: Hedström P and Swedberg R (eds.) *Social Mechanisms: An Analytical Approach to Social Theory*, 102–124. Cambridge, UK: Cambridge University Press.

Gangl M (2010) Causal inference in sociological research. *Annual Review of Sociology* 36: 21–47.

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

Gerring J (2007) Review article: the mechanistic world view: thinking inside the box. *British Journal of Political Science* 38: 161–179.

Gill S and Reynolds AJ (1999) Educational expectations and school achievement of urban African American children. *Journal of School Psychology* 37(4): 403–424.

Giuliani L (2003) Kleines Plädoyer für eine archäologische Hermeneutik, die nicht mehr verstehen will, als sie auch erklären kann, und die nur soviel erklärt, wie sie verstanden hat. In: Heinz M; Eggert MKH and Veit U (eds.) *Zwischen Erklären und Verstehen? Beiträge zu den erkenntnistheoretischen Grundlagen archäologischer Interpretation*, 9–22. Münster: Waxmann.

Glaesser J and Cooper B (2011) Selectivity and flexibility in the German secondary school system: a configurational analysis of recent data from the German Socio-Economic Panel. *European Sociological Review* 27(5): 570–585.

Glennan SS (1996) Mechanisms and the nature of causation. *Erkenntnis* 44(1): 49–71.

Glennan SS (2002) Rethinking mechanistic explanation. *Philosophy of Science* 69(S3): S342–S353.

Goffman E (1974) *Frame Analysis: An Essay on the Organization of Experience*. London: Harper and Row.

Goldberger AS and Manski CF (1995) Review article: The Bell Curve by Herrnstein and Murray. *Journal of Economic Literature* 33(2): 762–776.

Goldstein H (1995) *Multilevel Statistical Methods*. London: Hodder Arnold.

Goldthorpe JH (1996a) 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 (1996b) Problems of "meritocracy". In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 255–287. Boulder, CO / Oxford: Westview Press.

Goldthorpe JH (1998) Rational action theory for sociology. *British Journal of Sociology* 49(2): 167–192.

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

Goldthorpe JH (2003) The myth of education-based meritocracy. *New Economy* 10(4): 234–239.

Goldthorpe JH (2007) ”Cultural Capital”: Some Critical Observations. *Sociologica* 1(2): 1–23.

Goldthorpe JH and Jackson M (2005) Education-based meritocracy - the barriers to its realisation. *Paper presented at the Start-Up Workshop of the EDUC Research Theme of the 6th EU Framework Network of Excellence “Economic Change, Quality of Life & Social Cohesion (EQUALSOC)”, Dec. 2-3, 2005, Mannheim Centre for European Social Research* .

Goldthorpe JH and Jackson M (2008) ‘Education-based meritocracy: the barriers to its realization. In: Lareau A and Conley D (eds.) *Social Class: How Does it Work?*, 93–117. New York: Russell Sage Foundation Publication.

Gottfredson LS (1981) Circumscription and compromise: a developmental theory of occupational aspirations. *Journal of Counseling Psychology* 28(6): 545–579.

Gotthelf A (1976) Aristotle's conception of final causality. *The Review of Metaphysics* 30(2): 226–254.

Graham JW (2009) Missing data analysis: making it work in the real world. *Annual Review of Psychology* 60: 549–576.

Graham JW and Donaldson SI (1993) Evaluating interventions with differential attrition: The importance of nonresponse mechanisms and use of follow-up data. *Journal of Applied Psychology* 78(1): 119–128.

Granovetter M (1978) Threshold models of collective behavior. *American Journal of Sociology* 83(6): 1420–1443.

Granovetter M and Soong R (1983) Threshold models of diffusion and collective behavior. *Journal of Mathematical sociology* 9(3): 165–179.

Gresch C (2009) Empfehlungsstatus, Übergangsempfehlung und der Wechsel in die Sekundarstufe I: Bildungsentscheidungen und soziale Ungleichheit. *Bildungsentscheidungen. Sonderheft der Zeitschrift für Erziehungswissenschaften* 230–256.

Gross N (2009) A pragmatist theory of social mechanisms. *American Sociological Review* 74(3): 358–379.

Harker R and May SA (1993) Code and habitus: comparing the accounts of Bernstein and Bourdieu. *British Journal of Sociology of Education* 14(2): 169–178.

Hauser RM (1970a) Context and consex: a cautionary tale. *American Journal of Sociology* 75(4): 645–664.

Hauser RM (1970b) Reply to Barton. *The American Journal of Sociology* 76(3): 517–520.

Hauser RM (1974) Contextual analysis revisited. *Sociological Methods & Research* 2(3): 365–375.

Hauser RM (1976) On Boudon's model of social mobility. *American Journal of Sociology* 911–928.

Healy MJR and Goldstein H (1978) Regression to the mean. *Annals of human biology* 5(3): 277–280.

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

Heckman JJ (1995) Lessons from the bell curve. *Journal of Political Economy* 103(5): 1091–1120.

Hedström P (2005) *Dissecting The Social. On the Principles of Analytical Sociology.* Cambridge: Cambridge University Press.

Hedström P and Swedberg R (1996) Social mechanisms. *Acta Sociologica* 39(3): 281–308.

Hedström P and Swedberg R (1998a) *Social Mechanisms: An Analytical Approach to Social Theory.* Cambridge: Cambridge University Press.

Hedström P and Swedberg R (1998b) Social mechanisms: an introductory essay. In: Hedström P and Swedberg R (eds.) *Social Mechanisms: An Analytical Approach to Social Theory*, 1–31. Cambridge: Cambridge University Press.

Hedström P; Swedberg R and Udehn L (1998) Popper's situational analysis and contemporary sociology. *Philosophy of the Social Sciences* 28(3): 339–364.

Hedström P and Ylikoski P (2010) Causal mechanisms in the social sciences. *Annual Review of Sociology* 36: 49–67.

Heider F (1946) Attitudes and cognitive organization. *Journal of Psychology* 21(1): 107–112.

Hempel CG (1942) The function of general laws in history. *The Journal of Philosophy* 39(2): 35–48.

Hempel CG (1965) Aspects of scientific explanation. In: Hempel CG (ed.) *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, 331–496. New York: The Free Press.

Hempel CG and Oppenheim P (1948) Studies in the Logic of Explanation. *Philosophy of Science* 15(2): 135–175.

Herrnstein R and Murray C (1994) *The Bell Curve - Intelligence and Class Structure in America*. New York: The Free Press.

Hillmert S (2013) Links between immigration and social inequality in education: A comparison among five European countries. *Research in Social Stratification and Mobility* 32: 7–23.

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.

Hoffer TB (2001) Meritocracy. In: Levinson D; Cookson PW and Sadovnik AR (eds.) *Education and Sociology: An Encyclopedia*, 435–442. New York and London: Routledge.

Holland PW (1986) Statistics and causal inference. *Journal of the American statistical Association* 81(396): 945–960.

Horkheimer M and Adorno TW (2002) *Dialectic of Enlightenment: Philosophical Fragments*. Stanford: Stanford University Press.

Hox JJ (2010) *Multilevel Analysis - Techniques and Applications*. 2nd edn. New York and Hove: Routledge.

Hume D (1913) *An enquiry concerning the principles of morals (repr. from the ed. of 1777)*. Leipzig: Meiner.

Hyman H (1953) The Value systems of different classes: A social psychological contribution to the analysis of stratification. In: Bendix R and Lipset SM (eds.) *Class, Status and Power*, 426–442. Glencoe: The Free Press.

Jackson M; Erikson RS; Goldthorpe JH et al. (2007) Primary and secondary effects in class differentials in educational attainment. *Acta Sociologica* 50(3): 211–229.

Jæ ger MM and Holm A (2012) Conformists or rebels? Relative risk aversion, educational decisions and social class reproduction. *Rationality and Society* 24(2): 221–253.

Jaeger MM (2007) Economic and social returns to educational choices. *Rationality and Society* 19(4): 451–483.

Jalbert JE (2008) Husserl's position between Dilthey and the Windelband-Rickert school of Neo-Kantianism. *Journal of the History of Philosophy* 26(2): 279–296.

Jencks C (1972) *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

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

Jensen AR (1974) Ethnicity and scholastic achievement. *Psychological Reports* 34(2): 659–668.

Jensen AR (1986) g: artifact or reality? *Journal of Vocational Behavior* 29(3): 301–331.

Jensen AR (1987) The" g" beyond factor analysis. In: Glover JA; Conoley JC and Witt JC (eds.) *The Influence of Cognitive Psychology on Testing and Measurement*, 87–142. Hillsdale: Lawrence Erlbaum Associates, Inc.

Jensen AR and Reynolds CR (1982) Race, social class and ability patterns on the WISC-R. *Personality and Individual Differences* 3(4): 423–438.

Johnson VE (1996) On Bayesian analysis of multirater ordinal data: an application to automated essay grading. *Journal of the American Statistical Association* 91(433): 42–51.

Johnson VE (1997) An alternative to traditional GPA for evaluating student performance. *Statistical Science* 12(4): 251–269.

Jöreskog KG (1993) Testing structural equation models. In: Bollen KA and Long JS (eds.) *Testing Structural Equation Models*, 294–316. London: Sage Publications Ltd.

Jöreskog KG (1994) On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* 59(3): 381–389.

Jussim L (1986) Self-fulfilling prophecies: a theoretical and integrative review. *Psychological Review* 93(4): 429–445.

Jussim L (1989) Teacher expectations: self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology* 57(3): 469–480.

Jussim L; Eccles J and Madon S (1996) Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. In: Zanna MP (ed.) *Advances in Experimental Social Psychology*, 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.

Kahneman D and Tversky A (1984) Choices, values, and frames. *American Psychologist* 39(4): 341–350.

Kant I (1850) *Critique of Pure Reason.* London: Henry G. Bohn.

Kant I (1902) *Prolegomena to Any Future Metaphysics.* Chicago: Open Court Publishing Company.

Karlson KB and Holm A (2011) Decomposing primary and secondary effects: a new decomposition method. *Research in Social Stratification and Mobility* 29(2): 221–237.

Kelle U and Lüdemann C (1995) Grau, teurer Freund, ist alle Theorie... Rational Choice und das Problem der Brückenannahmen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47(2): 249–267.

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

Kemp S and Holmwood J (2003) Realism, regularity and social explanation. *Journal for the Theory of Social Behaviour* 33(2): 165–187.

Kendall MG (1951) *The Advanced Theory of Statistics*. New York: Hafner.

Kloosterman R; Ruiter S; De Graaf PM et al. (2009) Parental education, children's performance and the transition to higher secondary education: trends in primary and secondary effects over five Dutch school cohorts (1965–99). *The British Journal of Sociology* 60(2): 377–398.

Koertge N (1979) The methodological status of Popper's rationality principle. *Theory and Decision* 10(1): 83–95.

Korenman S and Winship C (2000) A reanalysis of The Bell Curve: intelligence, family background, and schooling. In: Arrow K; Bowles S and Durlauf S (eds.) *Meritocracy and Economic Inequality*, 137–178. Princeton, New Jersey: Princeton University Press.

Kozlowski SW and Kirsch MP (1987) The systematic distortion hypothesis, halo, and accuracy: an individual-level analysis. *Journal of Applied Psychology* 72(2): 252–261.

Kristen C (1999) *Bildungsentscheidungen und Bildungsungleichheit: Ein Überblick über den Forschungsstand*, vol. 5. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES).

Kroneberg C (2005) Die Definition der Situation und die variable Rationalität der Akteure: Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie* 34: 344–363.

Kroneberg C (2006) The definition of the situation and variable rationality: The model of frame selection as a general theory of action. *Sonderforschungsbereich 504 Publications, University of Mannheim* .

Kroneberg C (2011) *Die Erklärung sozialen Handelns. Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kroneberg C; Heintze I and Mehlkop G (2008) On shoplifting and tax fraud: An action-theoretic analysis of crime. *Sonderforschungsbereich 504 Publications, University of Mannheim* .

Kroneberg C and Kalter F (2012) Rational Choice Theory and Empirical Research: Methodological and Theoretical Contributions in Europe. *Annual Review of Sociology* 38: 73–92.

Kroneberg C; Yaish M and Stocké V (2010) Norms and rationality in electoral participation and in the rescue of Jews in WWII. *Rationality and Society* 22(1): 3–36.

Lakatos I (1978) *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*. Cambridge: Cambridge University Press.

Lareau A and Weininger EB (2003) Cultural capital in educational research: a critical assessment. *Theory and Society* 32(5): 567–606.

Lave CA and March JG (1975) *An Introduction to Models in the Social Sciences*. New York / Evanston / San Francisco / London: Harper and Row.

Lewis D (1973) Counterfactuals and comparative possibility. *Journal of Philosophical Logic* 2(4): 418–446.

Lewis D (1977) Possible-world semantics for counterfactual logics: a rejoinder. *Journal of Philosophical Logic* 6(1): 359–363.

Lewis D (1979) Counterfactual dependence and time's arrow. *Noûs* 13(4): 455–476.

Lewis D (1981) Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10(2): 217–234.

Lindenberg S (1985) An assessment of the new political economy: its potential for the social sciences and for sociology in particular. *Sociological Theory* 3(1): 99–114.

Lindenberg S (1990) Homo socio-oeconomicus: the emergence of a general model of man in the social sciences. *Journal of Institutional and Theoretical Economics* 146(4): 727–748.

Lindenberg S and Frey BS (1993) Alternatives, frames, and relative prices: a broader view of rational choice theory. *Acta Sociologica* 36(3): 191–205.

Linn RL (1973) Fair test use in selection. *Review of Educational Research* 43(2): 139–161.

Little D (1991) *Varieties of Social Explanation: An Introduction to the Philosophy of Science*. Boulder, CO: Westview Press.

Little RJA (1992) Regression with missing X's: a review. *Journal of the American Statistical Association* 87(420): 1227–1237.

Loehlin JC; Vandenberg SG and Osborne RT (1973) Blood group genes and Negro-White ability differences. *Behavior Genetics* 3(3): 263–270.

Lucas SR (2001) Effectively maintained Iinequality: education transitions, track mobility, and social background effects. *American Journal of Sociology* 106(6): 1642–1690.

Lucas SR (2009) Stratification theory, socioeconomic background, and educational attainment. *Rationality and Society* 21(4): 459–511.

Lucchini M; Della Bella S and Pisati M (2013) The weight of the genetic and environmental dimensions in the inter-generational transmission of educational success. *European Sociological Review* 29(2): 289–301.

Lüdtke O; Köller O; Marsh HW et al. (2005) Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology* 30(3): 263–285.

Luhmann N (1995) *Social Systems*. Stanford: Stanford University Press.

Lynch SM (2007) *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. New York: Springer.

Maaz K; Trautwein U; Lüdtke O et al. (2008) Educational transitions and differential learning environments: how explicit between-school tracking contributes to social inequality in educational outcomes. *Child Development Perspectives* 2(2): 99–106.

Maccoby EE (1992) The role of parents in the socialization of children: An historical overview. *Developmental Psychology* 28(6): 1006–1017.

Machamer P; Darden L and Craver CF (2000) Thinking about mechanisms. *Philosophy of Science* 67(1): 1–25.

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; Willard J; Guyll M et al. (2006) Self-fulfilling prophecy effects of mothers' beliefs on children's alcohol use: accumulation, dissipation, and stability over time. *Journal of personality and social psychology* 90(6): 911–926.

Mahoney J (2000) Strategies of causal inference in small-N analysis. *Sociological Methods & Research* 28(4): 387–424.

Mahoney J (2001) Review essay: beyond correlational analysis: recent innovations in theory and method. In: *Sociological Forum*, vol. 16, 575–593. Springer.

Makreel RA (1992) *Dilthey - Philosopher of the Human Studies*. Princeton, New Jersey: Princeton University Press.

Manski CF (1993) Identification problems in the social sciences. *Sociological methodology* 23(1993): 1–56.

Marsh HW (1987) The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology* 79(3): 280–295.

Marsh HW (1998) Pairwise deletion for missing data in structural equation models: non-positive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal* 5(1): 22–36.

Marsh HW; Chit-Kwong K and Kit-Tai H (2000) Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology* 78(2): 337–349.

Marsh HW; Köller O and Baumert J (2001) Reunification of East and West German school systems: longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal* 38(2): 321–350.

Marsh HW and Parker JW (1984) Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology* 47(1): 213–231.

Marsh HW; Seaton M; Trautwein U et al. (2008) The big-fish–little-pond-effect stands up to critical scrutiny: implications for theory, methodology, and future research. *Educational Psychology Review* 20(3): 319–350.

Maus I (2006) Der Urzustand. In: Höffe O (ed.) *Klassiker Auslegen: John Rawls - eine Theorie der Gerechtigkeit*, 71–95. Berlin: Akademie Verlag.

Maxwell JA (2004) Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher* 33(2): 3–11.

Mayntz R (2004) Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences* 34(2): 237–259.

McFarland C and Buehler R (1995) Collective self-esteem as a moderator of the frog-pond effect in reactions to performance feedback. *Journal of Personality and Social Psychology* 68(6): 1055–1070.

Mechtenberg L (2009) Cheap talk in the classroom: How biased grading at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies* 76(4): 1431–1459.

Menzies P (2009) Counterfactual Theories of Causation. In: *The Stanford Encyclopedia of Philosophy.*
URL: *http://plato.stanford.edu/entries/causation-counterfactual/*

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

Merton RK (1957) On sociological theories of the middle range. In: *On the Theoretical Sociology. Five Essays, Old and New*, 39–72. New York: The Free Press.

Merton RK (1987) Three fragments from a sociologist's notebooks: establishing the phenomenon, specified ignorance, and strategic research materials. *Annual Review of Sociology* 13.

Meulemann H (1976) *Wortbedeutungsverständnis und Wortbedeutungsexplikationen - Eine empirische Analyse zweier Aspekte des Sprachverhaltens und ihrer sozialen Determinanten im Rahmen der Theorie der linguistischen Codes. Materialien aus der Bildungsforschung Nr. 4.* Berlin: Max-Planck-Institut für Bildungsforschung.

Meulemann H (1979) Klassenlage, Entscheidungsfeld und Bildungsaspirationen - Ein Versuch zur theoretischen Präzisierung und kausalen Erklärung von Zusammenhängen zwischen sozialer Struktur und individueller Lebensplanung. *Zeitschrift für Soziologie* 8(4): 391–414.

Mill JS (1843) *The Logic of the Moral Sciences [1872].* London: John W. Parker, West Strand.

Mill JS (1863) *Die Methode der inductiven Forschung als die Methode der Naturforschung.* Braunschweig: Friedrich Vieweg und Sohn.

Mistry RS; White ES; Benner AD et al. (2009) A longitudinal study of the simultaneous influence of mothers' and teachers' educational expectations on low-income youth's academic achievement. *Journal of Youth and Adolescence* 38(6): 826–838.

Modin B; Erikson R and Vå gerö D (2013) Intergenerational Continuity in School Performance: Do Grandparents Matter? *European Sociological Review* (forthcoming).

Morgan SL (2006) Expectations and Aspirations. In: Ritzer G (ed.) *The Blackwell Encyclopedia of Sociology, Vol. 3*, 1528–1531. Malden, Mass: Wiley Blackwell.

Morgan SL (2011) Models of college entry in the United States and the challenges of estimating primary and secondary effects. *Cornell University Working Paper* .

Morgan SL and Winship C (2012) Bringing Context and Variability Back in to Causal Analysis. In: Kincaid H (ed.) *Oxford Handbook of the Philosophy of the Social Sciences*, chap. 14. Oxford: Oxford University Press.

Moynihan DP (1968) Sources of resistance to the Coleman Report. *Harvard Educational Review* 38(1): 23–36.

Muller C; Katz SR and Dance LJ (1999) Investing in teaching and learning: dynamics of the teacher-student relationship from each actor's perspective. *Urban Education* 34(3): 292–337.

Müller W and Karle W (1993) Social selection in educational systems in Europe. *European Sociological Review* 9(1): 1–23.

Müller-Benedict V (1999) Strukturelle Grenzen sozialer Mobilität. Ein Modell des Mikro-Makro-Übergangs nach Boudon. *Kölner Zeitschrift fur Soziologie und Sozialpsychologie* 51(2): 313–338.

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

Murphy KR; Jako RA and Anhalt RL (1993) Nature and consequences of halo error: a critical analysis. *Journal of Applied Psychology* 78(2): 218.

Muthén BO (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49(1): 115–132.

Muthén BO (1994) Multilevel covariance structure analysis. *Sociological Methods & Research* 22(3): 376–398.

Mutz DC (2002) The consequences of cross-cutting networks for political participation. *American Journal of Political Science* 46(4): 838–855.

Nadeau R (1993) Confuting Popper on the rationality principle. *Philosophy of the Social Sciences* 23(4): 446–467.

Nasco SA and Marsh KL (1999) Gaining control through counterfactual thinking. *Personality and Social Psychology Bulletin* 25(5): 557–569.

Nash R (2005) Boudon, realism, and the cognitive habitus: why an explanation of inequality / difference cannot be limited to a model of secondary effects. *Interchange* 36(3): 275–293.

Neal DA and Johnson WR (1996) *The Role of Pre-market Factors in Black-White Wage Differences.* Tech. Rep. 5, National Bureau of Economic Research.

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

Neugebauer M (2010) Bildungsungleichheit und Grundschulempfehlung beim Übergang auf das Gymnasium: Eine Dekomposition primärer und sekundärer Herkunftseffekte. *Zeitschrift für Soziologie* 39(3): 202–214.

Olsson UH (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 44(4): 443–460.

Opp KD (1991) DDR'89. Zu den Ursachen einer spontanen Revolution. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 43(2): 302–321.

Opp KD (2005a) Explanations by mechanisms in the social sciences. Problems, advantages and alternatives. *Mind & Society* 4(2): 163–178.

Opp KD (2005b) *Methodologie der Sozialwissenschaften: Einführung in Probleme ihrer Theorienbildung und praktischen Anwendung.* Wiesbaden: VS Verlag für Sozialwissenschaften.

Opp KD (2011) Modeling micro-macro relationships: problems and solutions. *Journal of Mathematical Sociology* 35: 209–234.

Pearl J (2000) *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

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.

Plato (1991) *The Republic.* New York: Vintage Books.

Plato (2009) *Meno.* Rockville, Maryland: Serenity Publishers.

Popper K (1959) *The Logic of Scientific Discovery.* London: Hutchinson.

Popper KR (1944a) The poverty of historicism, I. *Economica* 11(42): 86–103.

Popper KR (1944b) The poverty of historicism, II. *Economica* 11(43): 119–137.

Popper KR (1945a) *The Open Society and its Enemies.* London: Routledge.

Popper KR (1945b) The poverty of historicism, III. *Economica* 12(46): 69–89.

Popper KR (1957) *The Poverty of Historicism.* London: Routledge & Kegan Paul.

Popper KR (1963) *Conjectures and Refutations: The Growth of Scientific Knowledge.* London: Routledge.

Popper KR (1974) *Unended Quest.* La Salle, IL: Open Court Publishing & Co.

Popper KR (1985) The rationality princinple. In: Miller D (ed.) *Popper Selections*, 357–365. Princeton, New Jersey: Princeton University Press.

Popper KR (1994) Models, instruments, and truth: the status of the rationality principle in the social sciences. In: Notturno MA (ed.) *The Myth of the Framework.* New York: Routledge.

Quetelet A (1835) *Sur l'homme et le développement de ses facultés, ou Essai de physique sociale.* Paris: Bachelier.

Raftery AE and Hout M (1993) Maximally maintained inequality: expansion, reform, and opportunity in Irish education, 1921-75. *Sociology of Education* 66(1): 41–62.

Rasch G (1960) *Probabilistic models for some intelligence and attainment tests.* Koebenhavn: Danmarks paedagogiske Institut.

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.

Rawls J (1971) *A Theory of Justice.* Cambridge, Mass.: Belknap Press of Harvard University Press.

Rawls J (1974) Some reasons for the maximin criterion. *The American Economic Review* 64(2): 141–146.

Ready DD and Wright DL (2011) Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. *American Educational Research Journal* 48(2): 335–360.

Reskin BF (2003) Including mechanisms in our models of ascriptive inequality. *American Sociological Review* 68(1): 1–21.

Rheinberg F (1980) *Leistungsbewertung und Lernmotivation.* Göttingen: Hogrefe.

Rickert H (1899) *Kulturwissenschaft und Naturwissenschaft.* Freiburg: Mohr.

Rickert H (1902) *Die Grenzen der naturwissenschaftlichen Begriffsbildung - eine logische Einleitung in die historischen Wissenschaften.* Tübingen: Mohr.

Riedel M (1978) *Verstehen oder Erklären? Zur Theorie und Geschichte der hermeneutischen Wissenschaften.* Stuttgart: Klett-Cotta.

Riesman D (1967) Notes on meritocracy. *Daedalus* 96(3): 897–908.

Rigsby LC and McDill EL (1972) Adolescent peer influence processes: conceptualization and measurement. *Social Science Research* 1(3): 305–321.

Rindermann H and Heller KA (2005) The benefit of gifted classes and talent schools for developing students' competences and enhancing academic self-concept. *Zeitschrift für Pädagogische Psychologie* 19(3): 133–136.

Rindskopf D and Rose T (1988) Second order factor analysis: some theory and applications. *Multivariate Behavioral Research* 23(1): 51–67.

Ringer F (2002) Max Weber on causal analysis, interpretation, and comparison. *History and Theory* 41(2): 163–178.

Robinson WS (1950) Ecological Correlations and the behavior of individuals. *American Sociological Review* 15(3): 351–357.

Roese NJ (1994) The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology* 66(5): 805–818.

Rohlfing I (2012) *Case Studies and Causal Inference: An Integrative Framework*. Basingstoke: Palgrave Macmillan.

Rosenberg M (1979) *Conceiving the Self*. New York: Basic Books.

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

Rubin DB (1986) Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics* 4(1): 87–94.

Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: Wiley.

Rushton JP and Jensen AR (2005) Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law* 11(2): 235–294.

Salmon WC (1971) *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

Sarrazin T (2010) *Deutschland schafft sich ab. Wie wir unser Land aufs Spiel setzen*. 2nd edn. München: DVA.

Saunders P (1997) Social mobility in Britain: an empirical evaluation of two competing explanations. *Sociology* 31(2): 261–288.

Scarr S; Pakstis AJ; Katz SH et al. (1977) Absence of a relationship between degree of White ancestry and intellectual skills within a Black population. *Human Genetics* 39(1): 69–86.

Schafer JL (1997) *Analysis of Incomplete Multivariate Data*. Boca Raton / London / New York / Washington, DC: Chapman & Hall.

Schafer JL and Graham JW (2002) Missing data: our view of the state of the art. *Psychological methods* 7(2): 147–177.

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.

Schubert F and Becker R (2010) Social inequality of reading literacy - a longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pair wise matching procedure. *Research in Social Stratification and Mobility* 28(1): 109–133.

Seaton M; Marsh HW and Craven RG (2009) Earning its place as a pan-human theory: universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology* 101(2): 403–419.

Seaton M; Marsh HW and Craven RG (2010) Big-fish-little-pond effect: generalizability and moderation - two sides of the same coin. *American Educational Research Journal* 47(2): 390–433.

Sewell WH; Haller AO and Ohlendorf GW (1970) The educational and early occupational status attainment process: replication and revision. *American Sociological Review* 35(6): 1014–1027.

Sewell WH; Haller AO and Portes A (1969) The educational and early occupational attainment process. *American Sociological Review* 34(1): 82–92.

Simon HA (1955) A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1): 99–118.

Simon HA (1957) *Models of Man.* New York: Wiley.

Snijders TAB and Bosker RJ (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage Publications Ltd.

Solga H (2002) 'Stigmatization by negative selection' - explaining less-educated people's decreasing employment opportunities. *European Sociological Review* 18(2): 159–178.

Solga H (2005) Meritokratie - die moderne Legitimation ungleicher Bildungschancen. In: Berger PA and Kahlert H (eds.) *Institutionalisierte Ungleichheiten: Wie das Bildungswesen Chancen blockiert*, 19–38. Weinheim & München: Juventa.

Solga H and Becker R (2012) Soziologische Bildungsforschung - eine kritische Bestandsaufnahme. In: Becker R and Solga H (eds.) *Soziologische Bildungsforschung. Sonderheft 52 der Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 7–43. Wiesbaden: VS Verlag für Sozialwissenschaften.

Spearman C (1904) "General intelligence," objectively determined and measured. *The American Journal of Psychology* 15(2): 201–292.

Stigler GJ and Becker GS (1977) De gustibus non est disputandum. *The American Economic Review* 67(2): 76–90.

Stinchcombe AL (1991) The conditions of fruitfulness of theorizing about mechanisms in social science. *Philosophy of the Social Sciences* 21(3): 367–388.

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.

Stocké V (2009) Adaptivität oder Konformität? Die Bedeutung der Bezugsgruppe und die Leistungsrealität der Kinder auf die Entwicklung elterlicher Bildungsaspirationen am Ende Grundschulzeit. *Bildungsentscheidungen. Sonderheft der Zeitschrift für Erziehungswissenschaften* 12: 257–280.

Stocké V (2010) Der Beitrag der Theorie rationaler Entscheidung zur Erklärung von Bildungsungleichheit. In: Quenzel G and Hurrelmann K (eds.) *Bildungsverlierer - Neue Ungleichheiten*, 73–94. Wiesbaden: Springer.

Stocké V (2011) Educational Aspirations and Inequality in Educational Opportunity: The Difference between Realistic and Idealistic Aspirations. *Presentation at the RC28 Spring Meeting of the International Sociological Association (Essex, UK)* .

Stouffer SA (1950) *The American Soldier.* Princeton, New Jersey: Princeton University Press.

Sun R; Slusarz P and Terry C (2005) The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological review* 112(1): 159–192.

Tanner MA and Wong WH (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82: 528–540.

Thijs J; Verkuyten M and Helmond P (2010) A further examination of the big-fish-little-pond effect. *Sociology of Education* 83(4): 333–345.

Thomas WI and Thomas DS (1928) *The Child in America: Behavior Problems and Programs.* New York: Knopf.

Thorburn WM (1918) The myth of Occam's razor. *Mind* 27(3): 345–353.

Thorndike EL (1920) A constant error in psychological ratings. *Journal of Applied Psychology* 4(1): 25–29.

Thorndike RL (1971) Concepts of culture-fairness. *Journal of Educational Measurement* 8(2): 63–70.

Thorngate W (1980) Efficient decision heuristics. *Behavioral Science* 25(3): 219–225.

Tizard B (1974) IQ and Race. *Nature* 247(01): 316.

Tolsma J; Need A and de Jong U (2010) Explaining participation differentials in Dutch higher education: the impact of subjective success probabilities on level choice and field choice. *European Sociological Review* 26(2): 235–252.

Trautwein U and Lüdtke O (2005) The big-fish-little-pond effect. Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie* 19(3): 137–140.

Trautwein U; Lüdtke O; Marsh HW et al. (2006) Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology* 98(4): 788–806.

Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–1131.

Udehn L (2002) The changing face of methodological individualism. *Annual Review of Sociology* 28: 479–507.

University for Peace (ed.) (2004) *Environmental Degradation as a Cause of Conflict in Dafur*. Khartoum: University for Peace.

Van De Werfhorst HG and Van Tubergen F (2007) Ethnicity, schooling, and merit in the Netherlands. *Ethnicities* 7(3): 416–444.

Van den Eeden P (1992) Multilevel theory and the underspecification of multilevel models. *Quality & Quantity* 26(3): 307–322.

von Wright GH (1971) *Explanation and Understanding*. London: Routledge & Kegan Paul.

Šuber D (2010) Social science between Neo-Kantianism and Philosophy of Life: The cases of Weber, Simmel, and Mannheim. *Historical Perspectives on Erklären and Verstehen* 267–290.

Weber M (1922) *Wirtschaft und Gesellschaft. Grundriss der Sozialökonomik*. Tübingen: Mohr.

Weber M (1948) *From Max Weber: Essays in Sociology*. London: Routledge.

Weber M (1964) *The Theory of Social and Economic Organization*. New York: The Free Press.

Weber M (1968) *Economy and Society: An Outline of Interpretive Sociology*. New York: Bedminster Press.

Weber M (1978) *Economy and Society - An Outline of Interpretive Sociology (Vol. 1 & 2)*. Berkeley: University of California Press.

Weber M (1981) Some categories of interpretative sociology. *Sociological Quaterly* 22(2): 151–180.

Weber M (1985) *Wirtschaft und Gesellschaft - Grundriss der verstehenden Soziologie*. Tübingen: Mohr.

Weber M (2002) Economy and society: an outline of interpretive sociology. In: Woolsey Biggard N (ed.) *Readings in Economic Sociology*, 24–37. Oxford: Blackwell Publishing.

Weingartner S (2013) Hochkulturelle Praxis und Frame-Selektion. *Kölner Zeitschrift fur Soziologie und Sozialpsychologie* 65(1): 3–30.

Wilks SS (1932) Moments and distributions of estimates of population parameters from fragmentary samples. *The Annals of Mathematical Statistics* 3(3): 163–195.

Windelband W (1894) Geschichte und Naturwissenschaft. In: Windelband W (ed.) *Präludien*, 355–379. Tübingen: Mohr.

Winship C and Mare RD (1992) Models for sample selection bias. *Annual Review of Sociology* 18: 327–350.

Wippler R (1978) The structural-individualistic approach in Dutch sociology. *The Netherlands Journal of Sociology* 14: 135–155.

Woodward J (2002) What is a mechanism? A counterfactual account. *Philosophy of Science* 69(S3): 366–377.

Woodward J (2008) Causation and Manipulability. In: Zalta EN (ed.) *Stanford Encyclopedia of Philosophy*. Stanford: Center for the Study of Language and Information, Stanford University.
URL: *http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=causation-mani*

Wothke W (1993) Nonpositive definite matrices in structural modeling. *Testing Structural Equation Models* 154: 256–293.

Yaish M and Katz-Gerro T (2012) Disentangling 'cultural capital': the consequences of cultural and economic resources for taste and participation. *European Sociological Review* 28(2): 169–185.

Young M (1958) *The Rise of the Meritocracy, 1870-2033: An Essay on Education and Inequality*. London: Thames & Hudson.

Zeidner M and Schleyer EJ (1998) The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology* 24(4): 305–329.

II. Teachers' Evaluations and the Definition of the Situation in the Classroom

(with Klaus Birkelbach)

While a multitude of American, Swiss and German studies analyzed the predictors of kindergarten and primary school teachers' evaluations or transition recommendations, respectively, we did not find comparable evidence regarding teachers' evaluations measured mid of secondary school. In the theoretical section, we regard German 10^{th} class teachers' evaluations with a prognostic claim about students' future academic ability as a result of a special social situation in the classroom; and we synthesize existing meritocratic and *habitus*-related explanations to a more general theory of action according to the *M*odel of *F*rame *S*election (MFS). In the empirical section, we test both meritocratic and *habitus*-related hypotheses in a set of structural equation models. Using data from the *C*ologne *H*igh *S*chool *P*anel (*CHiSP*) we find that even when controlling for the model's path structure, indicators for both kinds of concepts are statistically significant. However, notwithstanding the underlying type of information processing, the predictive power of indicators operationalizing the meritocratic explanation is comparatively higher.

1 Introduction

The literature in educational sociology tends to agree in that teachers' expectations, judgments or evaluations can be an important dimension of social inequality in educational research. Regarding teachers' evaluations as a *treatment variable*, at least since the Wisconsin model of status attainment (Sewell et al., 1969, 1970), teachers' evaluations are acknowledged to play a decisive role in individuals' both educational and occupational status attainment process.

Regarding teachers' evaluations as an *outcome*, in the American context, a number of studies has shown that teachers' evaluations vary by students' social backgrounds, their ethnic group, and both social and racial student-teacher matches and mismatches (Alexander et al., 1987; Page, 1987; Farkas et al., 1990; Farkas, 2003; Ferguson, 2003;

Downey and Pribesh, 2004; Hughes et al., 2005; Ready and Wright, 2011).[1] In the German context, a great deal of literature has focused on the determinants of primary school teachers' transition recommendations regarding the secondary school track choice suggested to their students (Ingenkamp, 1971; Becker, 2003; Bos and Pietsch, 2004; Jürges and Schneider, 2006; Arnold et al., 2007; Ditton, 2007; Pietsch and Stubbe, 2007).

In both the German and the American case, researchers mainly focused on teachers' evaluations at students' early ages – being it in kindergarten or in primary school. Thus, little is known whether the determinants of teachers' evaluations found in these studies apply to teachers' evaluations at secondary school to the same extent. In particular, we aim to analyze German 10^{th} class *Gymnasium* teachers' evaluations regarding their students' prospective aptitude for academic studies, and regarding the latter transitions, the current state of the art is quite puzzling.

On the one hand, various theories propose that the effect of students' social backgrounds on their educational opportunities should decline both over time and in the course of students' educational transitions (Müller and Karle, 1993; Raftery and Hout, 1993; Breen and Goldthorpe, 1997; Lucas, 2001, 2009). On the other hand, several more recent studies also found social background effects to increase from upper secondary to tertiary education (Selz and Vallet, 2006; Erikson, 2007; Mayer et al., 2007; Müller and Pollak, 2007; Lörz and Schindler, 2009; **?**). Hence, even seemingly trivial questions such as whether students' social background effects on secondary school teachers' evaluations are still present after having controlled for student achievement indicators – which *is* observable for primary school teachers' recommendations – remain unanswered yet. As a second contradiction to the waning coefficients hypothesis, Andersen and Hansen (2012) deduced from Bourdieu's theory of cultural capital that the impact of its stylistic or symbolic components should increase with higher educational levels; and concerning students' educational performance, the authors find significant support for this claim. Therefore, we see need to test whether similar mechanisms are still present regarding secondary school teachers' evaluations.

A theoretical shortcoming of the literature about teachers' evaluations in general is that although proponents of rational-choice explanations of educational inequality have invested a great deal of effort in modeling expectations and cost-benefit evaluations of both students and their parents (Breen and Goldthorpe, 1997; Esser, 1999; Goldthorpe, 1996), action theories of teachers' assessments have not progressed with similar pace (Ditton, 2007). Therefore, as an additional, more conceptual aim, we intend to synthesize common hypotheses about how teachers' evaluations are shaped to a more abstract theory of action.

Our indicator of 10^{th} class teachers' evaluations is taken from the Cologne High School Panel (*CHiSP*), a German Panel dataset initially surveyed in 1969. In line with postulates by both early status attainment theorists (Sewell et al., 1969, 1970) and *Pygmalion* or self-fulfilling prophecy studies (Rosenthal and Jacobson, 1968; Madon et al., 1997;

[1] For a review of older (and in part less elaborate) studies analyzing the correlation between teachers' achievement judgments and students' actual achievement see Hoge and Coladarci (1989).

Jussim and Harber, 2005), it has been shown that these kind of evaluations actually have a direct impact on students' educational transitions (Becker, 2010) – which, in turn, affect long-term occupational positions via path dependencies (Birkelbach, 2011). Thus, another contribution of our paper lies in analyzing the conditions which shape evaluations of the "significant others" (Sewell et al., 1969) that directly or indirectly influence long-term educational and occupational status attainment processes.

This is carried out by regarding teachers' evaluations about students' future academic ability as a result of a specific *social situation* in the classroom. In the following theoretical section (section 2), we will first replicate the general model of sociological explanations as it has been introduced by Coleman (1990). Thereafter, in a brief reference to the *M*odel of *F*rame *S*election (MFS), we will discuss how existing meritocratic and *habitus*-related explanations of teachers' evaluations can be synthesized into the assumption of action scripts of a more automatic (*as-mode*) and a more rational (*rc-mode*) type of information processing, respectively (section 2.1).

After that we will summarize the state of the art in German and Swiss educational research about predictors of teachers' *recommendations* – supplemented by a review of several international studies regarding teachers' *evaluations* –, and we will deduce corresponding hypotheses for the specific kind of evaluations in our data (section 2.2). In section 3, we will briefly describe our data, indicators, and research design. Since we hypothesize a more complex path structure for some of our theoretical concepts, we will test our hypotheses via structural equation modeling (SEM). In section 4, we will discuss our main findings from our structural equation models. Most important, students' average grade is the strongest predictor in our models while intelligence comes second. This leads us to the conclusion (section 5) that the meritocracy explanation of teachers' evaluations – regardless of the underlying type of information processing – is empirically more pronounced than the explanation based on criteria conventionally associated with the efficacy of students' *habitus*. However, since we recognized several additional path coefficients for manifest variables that may be related to unconscious and automatic status-related mental processes often referred to as effects of *habitus*, we demand from further studies to develop a more elaborate measurement model of the underlying processes both on the teacher and on the student side than we were able to analyze with our data. Considering also teachers' backgrounds would then lead to a multilevel structural equation model with teachers' evaluations nested in both student- and teacher-level contexts.

2 Theory and Hypotheses

A general model of sociological explanations was given in Coleman's (1990) seminal monograph wherein he differentiates between macro-level and micro-level propositions as a general form of modeling individual behavior in specific social contexts. The three-step procedure from the macro-level to the micro-level and back to the macro-level was extended by Esser (1993, 1996, 1999) who labeled the steps as the *logic of the situation*, the *logic of selection*, and the *logic of aggregation*.

The logic of the situation describes the top-down link from the macro-level to the micro-level entailing assumptions about both the conditions of the social situation and the alternatives of individual actors. Actors' expectations and evaluations are linked to the conditions and alternatives of the social situation via *bridge hypotheses*.[2] The logic of selection aims to explain individual decisions on the micro-level based on an underlying theory of action. If the latter is described explicitly, scholars conventionally make use of *rational choice* (*RC*) or *subjective expected utility* (*SEU*) theory, or, as explicated below, of the *M*odel of *F*rame *S*election (MFS).

The logic of aggregation 'simply' embodies the bottom-up link between individual behavior on the micro-level and a collective explanandum on the macro-level via *transformation rules* that may vary depending on the respective context. Figure 1 displays this general scheme of sociological explanations.

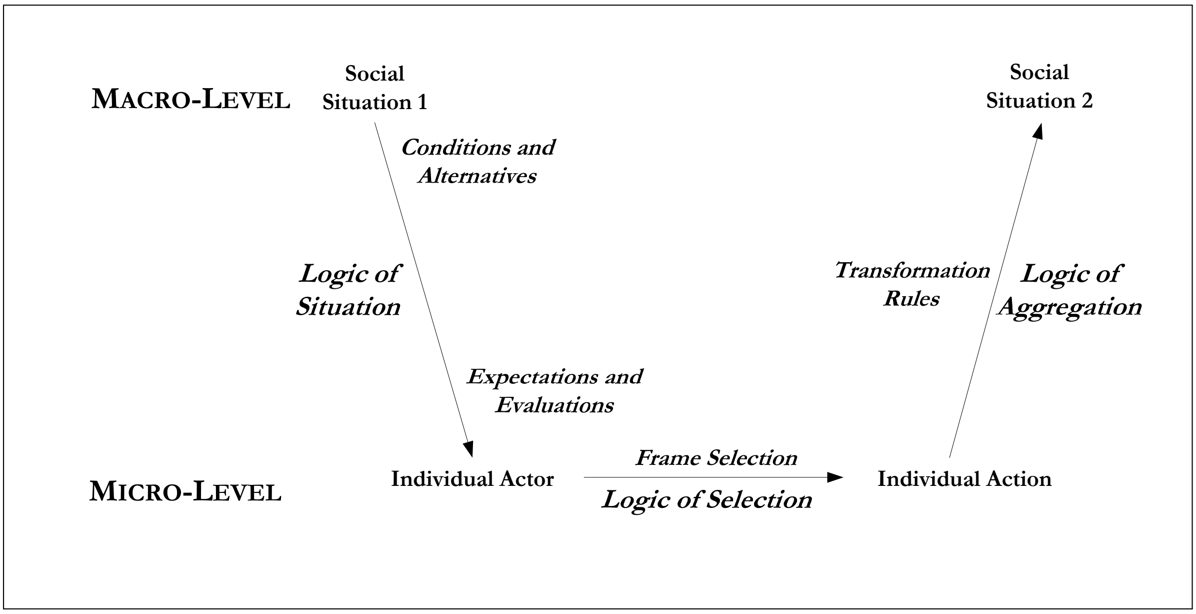


Figure 1: A general model of sociological explanations. Source: Esser (1993, p. 98).

Subject matter of our investigation is a specific form of teachers' evaluations concerning their students' ability for academic studies. In concrete terms, we refer to teachers' nominations whom of their students they consider to be able to start academic studies and, likewise, whom they consider to lack these prerequisites.[3] For an explanation of the emergence of these particular evaluations, a more detailed description of the social situation in the classroom is fruitful. Since teachers' evaluations will surely depend on their respective expectations of students' prospective achievement, our aim is to specify the

[2]Some authors also use the term 'bridge hypotheses' to refer to specifications of the degree of actors' rationality (e.g. Kelle and Lüdemann, 1995). We follow Kroneberg and Kalter (2012) in that we would these latter statements as *auxiliary assumptions* in order to distinguish them from the bridge assumptions associated with the logic of the situation.

[3]A more detailed description of our data and our dependent variable is given in section 3.

relevant bridge hypotheses that are necessary to link these expectations and evaluations to the conditions of the underlying social situation.

2.1 The Social Situation in the Classroom

In the literature on teachers' *recommendations*, it is assumed that the latter are actually based on rational decisions and a 'correct' definition of the situation in order to ensure these recommendations to be somehow optimal for the students (Ditton, 2007).[4] In this sense, a 'correct' definition of the situation should be shaped by the idea of meritocracy, and it should consider both students' actual achievement and their future development possibilities. In the words of David Bell (1972, p. 41f.), "meritocracy is (...) the displacement of one principle of stratification by another, of ascription by achievement", and it is this idea that is supposed to legitimize for the selective function of the educational system (Solga, 2005).[5] Hence, we should keep in mind that when talking about 'rational' recommendations of the teachers, we always imply a kind of Weberian *ideal type* (Weber, 1968, p. 19-22) of objectivity and rationality that might (and *idealiter* also should) thoroughly serve as a frame for teachers' recommendations. However, all actual recommendations will never be more than a subjective and thus more or less imperfect realization of this ideal type of rationality.

In many – not all – German federal states ('Bundesländer'), teachers' recommendations concerning the transition from primary to the three-tiered secondary school (*Hauptschule*, *Realschule*, and *Gymnasium*) are legally binding.[6] Because of the minor permeability between lower and higher education within the stratified German school system, there are only small chances to adjust an inaccurate (but nonetheless binding) recommendation of the teacher or an inadequate parental transition decision during students' future educational life course (see e.g. Glaesser and Cooper, 2011). Actually, teachers' recommendations are more or less accurate forecasts of students' future achievement. One can expect them to be grounded on an evaluation of their actual

[4] Apart from Germany, we only know of the Netherlands and Switzerland as countries wherein students receive an explicit recommendation by their primary school teachers with regard to secondary school choice. In the Dutch case, teachers' recommendations are strongly influenced by students' results in a (compulsory) national achievement test (Tolsma et al., 2010). Both criteria were introduced in 1968 (in course of the Mammoth Law) in order to achieve a more meritocratic school system (Dronkers, 1993). Several studies noted that, possibly due to a success of all integrative ambitions, the effect of students' social backgrounds on teachers' recommendations decreased over time (for a review see Dronkers, 1983). In the Swiss canton Fribourg, transition recommendations at the end of primary school (i.e. in 6^{th} grade) have been introduced in the course of broader educational reforms mid of the 1990s in order to foster transparency and to even out educational inequalities (Baeriswyl et al., 2006, p. 378f.). Analyses based on student data from Fribourg indicate that in contrast to Germany, the effect of students' social backgrounds is partialed out when controls for achievement go into the model (Baeriswyl et al., 2006; Trautwein and Baeriswyl, 2007).

[5] For theoretical and empirical critiques of the meritocratic argument see Kaplan and Kaplan (1997); Breen and Goldthorpe (1999); Goldthorpe (2003); Goldthorpe and Jackson (2008) as well as Becker and Hadjar (2011).

[6] For a more detailed description of the German educational system see Hillmert and Jacob (2010); Jürges and Schneider (2006) as well as Pietsch and Stubbe (2007).

performance, and on additional information about familial endorsement even spanning students' prospective educational transitions. Thus, teachers' recommendations have far reaching consequences for students' further life course.

However, with regard to teachers' *evaluations* in our data, which are – in contrast to teachers' *recommendations* – neither made public to the students nor have a binding character for them, an explanation based on a too narrow notion of rationality may fall too short. One major reason is that these subjective evaluations lack any dependence on structural necessities of the school system – meaning that teachers' subjective assessments of students' academic ability will neither be influenced by assumptions about their direct impact on students' transition decisions nor by outright norms of the respective school environment.

In the American context, several studies argued that although even non-binding evaluations are relatively unbiased regarding students' actual performance (Hoge and Coladarci, 1989; Ready and Wright, 2011), nonetheless, cultural differences between students and their teachers might be an important factor in explaining teachers' evaluations apart from achievement-related criteria (Alexander et al., 1987; Farkas et al., 1990; Farkas, 2003; Ferguson, 2003; Downey and Pribesh, 2004; Morris, 2005). While in the latter studies, these cultural differences were largely related to ethnic group membership, in case of our data (German *Gymnasium* students surveyed in 1969), ethnic or racial issues are unlikely to be a dominant factor. Yet there are arguments that even for the specific kind of evaluations surveyed in our data, both achievement-related criteria and students' social backgrounds might shape teachers' judgments. In particular, we strive to provide a reconstruction of teachers' reasoning when evaluating their students by referring to Esser's and Kroneberg's enhancement of Kahneman and Tversky's (1984) early proposition of the framing approach towards a general theory of action (Esser, 1996, 2010; Esser and Kroneberg, 2010; Kroneberg, 2006; Kroneberg et al., 2008, 2010; Kroneberg, 2011; Kroneberg and Kalter, 2012).

Concretely, a particular frame defining an actor's social situation is usually given in terms of an automatic-spontaneous mode (*as-mode*), but within each such frame, certain *scripts* of action (and also actions themselves) may vary between a more habitual *as-mode* and a more rational *rc-mode* depending on the definitional complexity of the social situation.[7] In the *C*ologne *Hi*gh *S*chool *P*anel (*CHiSP*), for the teachers, the given *frame* of a survey-based assessment of students' prospective aptitude for academic studies should be relatively doubtless (*as-mode*).

Regarding the more concrete *scripts* of action, one the one hand, the most obvious (i.e. *as-mode*) script would be one that is consistent with the demand for professional pedagogic diagnostics according to meritocratic criteria such as students' academic performance. But in addition, several authors have argued that teachers can also be unconsciously influenced by students' *habitus* (Bourdieu, 1986; Bourdieu and Passeron, 1990) in terms of schemes of perception as well as of communication and self-control strate-

[7]More precisely, the probability of an *as-mode* script to be activated is positively associated with its general *availability*, its *accessibility* given the selection of a particular frame, and the *match* of that frame to the social situation at hand (Kroneberg, 2006, p. 18; also see Kroneberg, 2011, p. 132).

gies (De Graaf and De Graaf, 2002; De Graaf et al., 2000; De Graaf, 1986; DiMaggio, 1982; Dumais, 2006; Jaeger, 2009). Depending on each script's availability (Kroneberg, 2006, p. 18), a dominant *as-mode* script will approximate more to one of either types of meritocracy and *habitus*. While taking up the general assumption of actors' *variable rationality* (Kroneberg, 2005, 2006, 2011, 104-108), and in that case the idea of unconscious status-related mental processes, we would refrain from adhering to all ideological implications the concept of *habitus* has been revealed to comprise (cf. Lareau and Weininger, 2003 and Goldthorpe, 2007).

On the other hand, for both meritocracy- and *habitus*-related scripts, teachers might also find entirely *rational* justifications why they ground their evaluations on certain criteria (*rc-mode*). Concerning meritocracy, teachers could reflect that a student's current school performance does only imperfectly correspond to her actual cognitive capability and/or motivation that might nonetheless enable her to start academic studies at university. Regarding students' social backgrounds, though of course not being in line with the paradigm of meritocratic pedagogic diagnostics, teachers might hold the view that service-class students might be more successful at university than working-class students in that their parents, say having an academic background themselves, would be more able to support them, or that they better match to the *habitus* of their university teachers.[8]

In sum, we assume that in most cases, teachers ground their evaluations on meritocratic criteria according to a relatively automatic *as-mode* processing type – which can be distorted by unconscious status-related mental processes conventionally denoted as *habitus* effects. But the other hand, teachers may also refer to additional rational justifications in terms of both meritocracy- and status-related characteristics.[9]

2.2 Determinants of Teachers' Evaluations

Academic performance Being perhaps the most visible criterion, the predictive validity of school grades as the most common indicator of students' academic performance is, as several meta-analyses suggest, well-corroborated (Burton and Ramist, 2001; Kuncel et al., 2001; Morgan, 1989; Robbins et al., 2004). Although in Germany the value of school grades for long-term recommendations has been discussed since the 1920s (cf. Ingenkamp, 1971; Ziegenspeck and Lehmann, 1999), the average grades given by differ-

[8]Elster (1983, pp. 69-71, 101-108) and Hedström (2005, p. 4) are quite critical about the analytic precision of Bourdieu's notion of *habitus*. Collet (2009) and Yaish and Katz-Gerro (2012) are dissatisfied with a reduction of *habitus* on a merely unconscious, stimulus-response type of behavior. In line with the idea of variable rationality, we assume that teachers may also rationally reflect about status-related mental processes (*habitus-rc-mode*).

[9]Note that with the current data at hand, we are not able to test whether a given action script of either theoretical concept is actually more close to the *as-mode* or more close to the *rc-mode* of information processing. This is why the theoretical remarks above should be read more as a theoretical elucidation than in terms of 'rigor' groundwork whereof testable hypotheses would be deduced. However, in the conclusion section we will provide practical advice how empirical analyses might further proceed to test the prevalence of either action script reliant on teacher background variables and/or in a more comparative framework.

ent teachers over a longer time span are at least a good predictor of students' future academic success (Trapmann et al., 2007). Moreover, Arnold et al. (2007, p. 283) found that students' grades in mathematics and German language taken together can account for about two thirds of the total variance of teachers' recommendations. For both Germany (Tiedemann and Billmann-Mahecha, 2007; Gröhlich and Guill, 2009; Milek et al., 2009) and Switzerland (Trautwein and Baeriswyl, 2007), primary school teachers' transition recommendations were observed to be significantly predicted by students' grades even after controlling for class-level reference-group effects. Since the former findings are supported by American studies analyzing teachers' evaluations (Hughes et al., 2005; Ready and Wright, 2011)[10], and teachers' consideration of students' academic performance can be expected to be their probably most dominant (*as-mode*) script of action, we hypothesize:

H_1: *The better students' school grades, the higher their probability of obtaining a better evaluation.*

Cognitive ability According to Ingenkamp (1971), in the field of transition from primary to secondary school, test results have always been used to compensate for the fallibility of teachers' assessments.[11] In terms of predictive validity, also more recent studies highlight that standardized test scores would be more valid indicators than students' school grades (Camara, 1998; Camara and Echternacht, 2000; Camara et al., 2003).

Admittedly, cognitive capabilities can be regarded as the most important predictor of school achievement, but a considerable empirical gap between test results and *teachers'* evaluations can be detected notwithstanding: In their investigation of German primary school teachers' transition recommendations, Arnold et al. (2007, p. 281) found that students' reading literacy could account for only 31% of the variance of the teachers' recommendations.

Nevertheless, a linear relationship between intelligence and the probability of obtaining a particular teacher's *recommendation* to attend *Gymnasium* still holds – especially for the verbal component of intelligence (Ditton, 2007). Moreover, both Tiedemann and Billmann-Mahecha (2007) as well as Trautwein and Baeriswyl (2007) showed that apart from grades, test results still affect teachers' transition recommendations in Germany and Switzerland – even when the models are controlled for class-level reference-group effects. While in the American context, most studies used students' test results as a measure of their academic performance (see previous paragraph), Farkas et al. (1990) indeed found teachers' evaluations to be considerably correlated with students' test results (while

[10] Alexander et al. (1987) and Farkas et al. (1990) analyzed how teachers' evaluations serve as a mediating variable between students' social backgrounds, their cognitive abilities, and their school grades.

[11] Moede et al. (1919) and Bobertag and Hylla (1926) can be quoted as very early references for attempting to ground teachers' recommendations about school transition on the ground of standardized test results.

teacher-assigned grades were measured separately).[12] As we argued in the theoretical section, apart from students' academic performance, teachers might additionally try to estimate students' cognitive ability in order to rationally (*rc-mode*) increase the validity of their forecasts with regard to students' potential academic success at university. Thus we hypothesize:

H_2: The higher students' intelligence, the higher their probability of obtaining a better evaluation.

Social backgrounds Although the impact of students' social background variables on their school achievement is basically undoubted, both strength and importance of this relationship is still discussed broadly (Becker, 2003; Becker and Hecken, 2009; Shavit and Blossfeld, 1993; Breen and Goldthorpe, 1997; Breen and Jonsson, 2000; Breen et al., 2009, 2010; Erikson et al., 2005; Goldthorpe, 2003; Hillmert and Jacob, 2010; Schneider, 2008; Schubert and Becker, 2010; Stocké, 2007; Tolsma et al., 2010).

In general, the literature distinguishes between *primary effects* of social inequality which denote the impact of parental socioeconomic status (SES) on differences in students' academic abilities, and *secondary effects* of social inequality that capture differences – primarily in educational aspirations – *apart from* actual differences in academic abilities (Boudon, 1974).

As regards *primary* effects, Arnold et al. (2007, p. 287) also observed that the odds to attend *Gymnasium* is almost four times higher for "higher service class" children compared to "working class" children. Having controlled for students' cognitive abilities and reading literacy, these odds ratio reduces to 2.6. Hence, the difference between the two values can be interpreted as the result of primary effects of social inequality regarding teachers' formation of their evaluations due to students' actual ability differences (for similar results see Bos and Pietsch, 2004; Jürges and Schneider, 2006; Pietsch and Stubbe, 2007; Tiedemann and Billmann-Mahecha, 2007 and Gröhlich and Guill, 2009).[13] In American studies, significant effects of parental SES on teachers' evaluations are reported by Alexander et al. (1987); Farkas et al. (1990); Downey and Pribesh (2004); Hughes et al. (2005) and Ready and Wright (2011). These findings and the mechanisms discussed in our theoretical section provide us with good reasons to test for the supposition that parental SES might also influence teachers' evaluations as measured in our data:

H_3: The higher the socioeconomic status (SES) of students' parents, the higher their

[12] However, although Farkas et al. (1990) provided a variety of regression models, no regression model of the effect of test results on teachers' evaluations *net of* other predictors was reported. Also note that Farkas et al. (1990) used teachers' evaluations to operationalize students' *habits and styles* according to Bourdieu's (1986) notion of cultural capital – which does not allow to estimate the effect of *habitus*-related processes on teachers' evaluations (if not measured in terms of grades).

[13] As mentioned above, in the Swiss study by Trautwein and Baeriswyl (2007), the effect of parental SES on teachers' transition recommendations is canceled out when analyses control for students' grades, their motivation and their cognitive capabilities.

probability of obtaining a better evaluation.

As regards *secondary* effects, the "the inadequacy of uni-factorial theories" has been prominently criticized (Boudon, 1974, p. 101). The crucial point of this critique about merely one-factorial theories is that secondary effects of social inequality are still present after having controlled for all primary effects. That is, regardless of differences in cognitive abilities, "working class" children will still do less successfully in school because of lower educational expectations and aspirations.[14]

Our assumption is that students' aspirations not only impact their educational transitions – but also, previously, their teachers' evaluations that might thoroughly have an influence on the later transition decisions. The claim that this effect takes place independently of academic performance, cognitive abilities and even parental SES implies that students' aspirations somehow affect teachers' internalized norms and habits.

Several authors made use of students' aspirations to measure *habitus*-related components (McClelland, 1990; Dumais, 2002, 2006; Andres, 2009)[15], and also Morgan (2002, p. 423) argues in favor of using aspirations in terms of an "anticipation, based upon the unconscious estimation of the objective probabilities of success" (Bourdieu, 1973, p. 83). As outlined in section 2.1, if teachers have internalized certain norms and habits quite strongly, the latter might automatically enter teachers' dominant script of action in terms of an *as-mode* type of processing.

However, maintaining the idea of *variable rationality* as proposed by the MFS, teachers could also find rather rational arguments why students with certain social backgrounds in general and certain aspirations in particular might do better (*rc-mode*).

In sum, on a conceptual level, effects of students' aspirations on their educational transitions are explained as a form of *secondary* effects of social inequality in educational opportunities. Regarding teachers' evaluations as an outcome, on a processual level, we broaden the limited focus of the literature on merely habitual explanations by allowing for teachers' rational reflections on students' aspirations as well.[16]

[14] Given education as an investment good (Goldthorpe, 1996, p. 494), the chief concern for each family will be to achieve some kind of *inter-generational stability of class positions*. Hence, service-class parents will be more likely than others to encourage their children to attain some kind of higher education (Breen and Goldthorpe, 1997). 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). Thus, a higher level of education will be aspired if the educational motivation to continue somehow exceeds the underlying investment risk (also see Esser, 1999, pp. 265-275).

[15] Contrarily, van de Werfhorst and Hofstede (2007) are quite skeptic about the relation of students' *habitus* to their aspirations, and the authors did not find a significant effect in their regression models. However, van de Werfhorst and Hofstede (2007) merely operationalized *habitus* in terms of parental 'high brow' culture participation, and following a recent critique by Andersen and Hansen (2012), this is a too narrow understanding of the concept since students' working habits and their visible educational effort would be excluded (also see Farkas et al., 1990).

[16] Once more we have to stress that when we make use of 'habitual' explanations, we do not buy all implications associated with Bourdieu's concept of *habitus*. For instance, rational action theories on social inequality in educational opportunities dissect the latter into stratum-specific differences

Hence, our last hypothesis reads ass follows:

H_4: The higher students' aspirations, the higher their probability of obtaining a better evaluation.

3 Research Design

3.1 Data

All analyses will rely on a dataset which is known as the "*C*ologne *Hi*gh *S*chool *P*anel" (*CHiSP*). The *CHiSP* consists of an initial survey from 1969 with N=3385 10^{th} grade high school (*Gymnasium*) students in North Rhine-Westphalia and three re-surveys in 1985 (N=1987), 1996/97 (N=1596), and 2010 (N=1301). In the initial survey, students have been asked about issues like their performance, interests and plans in school, about their social origin and their relationship to their parents. Simultaneously to the initial survey, the students took part in an *I*ntelligence *S*tructure *T*est (IST) which consists of four sub-scales developed by Amthauer (1957). At the same time, also the students' teachers (N=1701) and parents (N=2646) have been surveyed. The main items of the parent questionnaire covered issues such as their social background, their child-raising practices, and their educational and occupational aspirations for their children.[17]

3.2 Variables

Dependent variable In the *CHiSP*, teachers have been asked to evaluate by a dichotomous decision whom of their students they suppose to be able for academic studies, and whom of them not. Since this was an open-ended question, teachers could classify students as being able, as 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

concerning parents' or students' underlying cost-benefit considerations (Boudon, 1974, p. 29ff.; Goldthorpe, 1996; Erikson and Jonsson, 1996; Breen and Goldthorpe, 1997; Esser, 1999, p. 265-275). While both in the latter studies as well as in the paper at hand, the less parsimonious assumption of variation in the absolute value of education *per se* is abandoned, exactly this is what cultural capital theorists in the tradition of Bourdieu (1986) maintain (McClelland, 1990; Dumais, 2002, 2006; Andres, 2009).

[17] In the first re-survey in 1985, the former students – at that time about 30 years old – provided detailed information about their private backgrounds and occupational careers beginning at the age of 15 until the age of 30. In the second re-survey in 1996/97, the time segment from the age of 30 until the age of 43 was added to the data, and in the third re-survey in 2010, the time segment until the age of 55 followed. Apart from the former students' life courses, common focus of the questionnaires were items about their biographical self-definition and -reflection, causal attribution, about the essential role of particular areas of life, and about their attitudes towards family, work and politics. For a general overview on the existing literature with the *CHiSP* data hitherto see Birkelbach (1998) and Meulemann et al. (2001).

analysis of the intra-class correlations (ICC) revealed a considerable variance of multiple teachers' evaluations for each student (see Becker and Birkelbach, 2010). Second, the question's openness might induce additional complication, because it has to be clarified whether the 'missing' category should really be considered as a missing value in statistical terms – or if we were to lose substantial information when proceeding on this assumption.

To overcome the first problem, our analyses will focus on evaluations only of class teachers.[18] 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, respectively, on the same independent variables which we will use in our structural equation models. These results are displayed in the appendix (tables B and C). We can note that for the analysis of the chance of getting a positive evaluation vs. getting none at all (table C), the effect sizes of all independent variables are of the same sign, but notably lower than for the analysis of the chance of getting a positive evaluation vs. getting a negative one (table B). This is also reflected in the explained variances of the two models, which are notably higher for the models in table B than for those in table C. Thus, we can conclude that students who have not been mentioned at all rank *lower* in their teachers' perceptions than students with a positive teacher evaluation, but *higher* than students with a negative teacher evaluation. To get to the point: When teachers do not receive clear evidence for their decision, they will shape only vague expectations for their students. Thus, in the subsequent structural equation models we will treat the 'missing' category not as missing – but as being located between teachers' positive and negative evaluations. Hence, our outcome will be coded as follows: 1 'not able'; 2 'not mentioned'; 3 'able'.

Independent variables First, students' *intelligence* was measured by their scores in an *I*ntelligence *S*tructure *T*est (Amthauer, 1957) consisting of four sub-scales each reflecting a distinct cognitive dimension (analogy, selection of words, series of numbers, cube test). For the structural equation models we will use the z-transformed scores of these sub-scales as a measure for the latent variable of students' intelligence (reflective indicators; see Bollen and Lennox, 1991; MacCallum and Browne, 1993; Diamantopoulos and Winklhofer, 2001).

Second, we control for students' *academic performance* in terms of their average grades.[19] Third, *parental socioeconomic status* (SES) will be operationalized as the maximum value of both mother's and father's education and occupational prestige (which equals the "dominance" approach suggested by Erikson, 1984). Education was mea-

[18] 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 other teachers – which is why their evaluations should be less error-prone (Raudenbush, 1984, p. 91). Thus, regarding only class teachers' evaluations will both simplify the data structure and provide a lower-bound estimate in particular of the 'less legitimate' predictors of teachers' evaluations.

[19] Note that according to the German grade system, an average grade *below* the median displays relatively *better* marks and an average grade *above* the median relatively *worse* marks. To ensure that higher variable values are associated with better marks, we inverted the variable.

sured in twelve categories ranking from 1 'without graduation' to 12 'university degree'. We categorized this variable into four dimensions (1 'lower'; 2 'middle'; 3 'higher'; 4 'degree'). Concerning occupational prestige, the data already contains the respective Treiman prestige scores (Treiman, 1977).[20]

Finally, students' aspirations are measured by their appraisement whether 'Abitur' is necessary to reach their aim in life – if any (1 'necessary'; 2 'not necessary, but useful'; 3 'not necessary'). We dichotomized this variable into 0 'no aim in life / Abitur not necessary'; 1 'Abitur useful or necessary'.[21]

3.3 Preliminary Path Structure and Plan of Analysis

Since we expect the independent variables to be correlated with each other considerably, we intend to model these intercorrelations directly in our estimations. We expect, first, that students' intelligence will be able to explain part of the variance of their school grades. Second, our considerations about the primary effect of social inequality (Boudon, 1974) imply that parental SES will influence both students' intelligence and their school grades. Third, to consider also the secondary effect of social inequality, we assume an impact of parental SES on students' aspirations. Fourth, it seems reasonable that higher grades will foster students' aspirations – and reversely. Therefore, we will allow for a covariance between those two variables. And finally, research about both *Pygmalion* and self-fulfilling prophecies[22] has shown that understanding teachers' evaluations as an entirely endogenous variables would fall too short. This is why we will model the relationship between school grades and teachers' evaluations and the one between students' aspirations and teachers' evaluations as covariances rather than as regression weights. The preliminary path model is presented in figure 2.

[20]See Ganzeboom and Treiman (1996) for a general overview about classification of occupations. Another possibility of dealing with parental SES would be to model all available information, i.e. all four variables, as formative indicators of a latent variable 'SES' (Bollen and Lennox, 1991; MacCallum and Browne, 1993; Diamantopoulos and Winklhofer, 2001). However, since the initial survey of the *CHiSP* took place in 1969, we have to expect that a considerable amount of mothers would not have been employed; hence, the variance of this variable would be rather low. Indeed, a brief glance at the distribution of occupational prestige by gender revealed that an amount of 78% of all mothers had not been in labour when they have been surveyed (not shown, available upon request). As a consequence, the factor loadings of a confirmatory factor analysis wherein we treated the four SES variables as formative indicators were rather low (not shown, available upon request). Thus, we consider introducing the maximum value of both mothers' and fathers' education and occupational prestige as two single indicators to be a better strategy and to lead to more consistent estimates.

[21]Table A (appendix) lists minimum/maximum, mean and standard deviation of all variables.

[22]For the initial study of *Pygmalion in the Classroom* see Rosenthal and Jacobson (1968). For early meta-analyses of existing studies about *Pygmalion* up to that point see Smith (1980) and Raudenbush (1984). For a more recent summary of implications and open questions in self-fulfilling prophecy research see Jussim and Harber (2005).

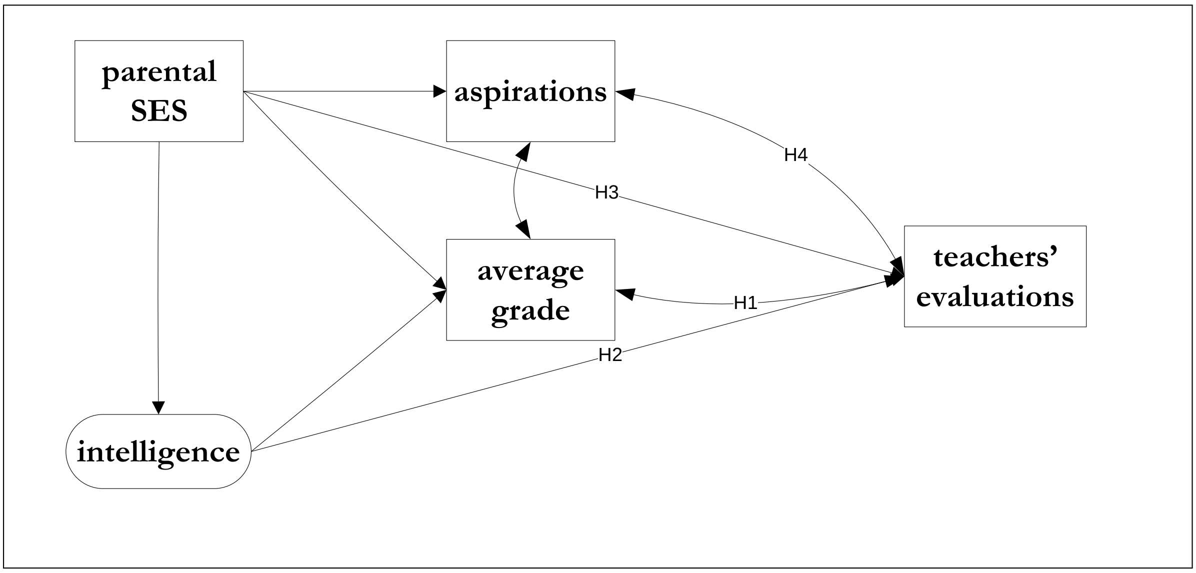


Figure 2: Preliminary path model.

3.4 Statistical Approach: Structural Equation Modeling

In order to take the complex path structure of the independent variables into account, we ran a set of structural equation models.[23] Since our dependent variable is categorical, conventional maximum likelihood estimation based on a usual variance-covariance matrix will be biased (Bollen, 1989, p. 433ff). Instead, it has been suggested to use a matrix of polychoric correlations (Olsson, 1979; Muthén, 1984; Aish and Jöreskog, 1990; Jöreskog, 1994) as input matrix.[24] The basic idea of polychoric correlations of categorical variables is to compute the thresholds of an hypothesized underlying continuous variable. To get a comparable metric for all variables, we also categorized the ratio-scaled variables in the data.[25] For our model, we have dichotomized the IST sub-scores, students' average grade and parental occupational prestige based on their respective median. The polychoric correlation matrix is displayed in table D (appendix). We used the `SEM` package in *R* (Fox, 2006) for our analyses.

[23]The SEM approach is also known as a LISREL model (Jöreskog and Sörbom, 1989; Jöreskog, 1993), named after the first statistical package which could deal with SEMs. Bollen (1989) is still the classical textbook for structure equation models.

[24]Maximum-Likelihood estimation of SEM models based on polychoric correlations may lead to consistent estimates, but the standard errors, z-values and significance parameters will be biased (Bollen, 1989, p. 443). Therefore, we use bootstrapping techniques to correct the latter parameters (Zhang and Browne, 2006; Fox, 2006).

[25]See Babakus et al. (1987) and Rigdon and Ferguson Jr (1991) for issues of convergence rates and fit statistics of polychoric correlations depending on different types of categorization.

4 Results

4.1 Measurement Part

Following the "Jöreskog tradition" (Byrne, 2004) in structural equation modeling, first of all the measurement model for the intelligence sub-scores was fitted (figure 3).[26]
The reflective measurement model for the intelligence scores (IST) achieved a good fit with respect to the *A*djusted *G*eneral *F*it *I*ndex ($AGFI = .996$), the *C*omparative *F*it *I*ndex ($CFI = .992$), the *R*oot *M*ean *S*quare *E*rror of *A*pproximation ($RMSEA = 0.018$) and the *S*tandardized *R*oot *M*ean *S*quare *R*esidual ($SRMR = 0.008$).[27]

The insignificant χ^2-value of 4.226 (df=2) suggests that there is no significant difference between the variance-covariance matrix of the observed variables and the model we have estimated. Looking at the standardized estimates, we can observe that all except one IST sub-dimensions show factor loadings around .45 – .50. Only the cube test seems to perform slightly worse in explaining the latent variable "intelligence".

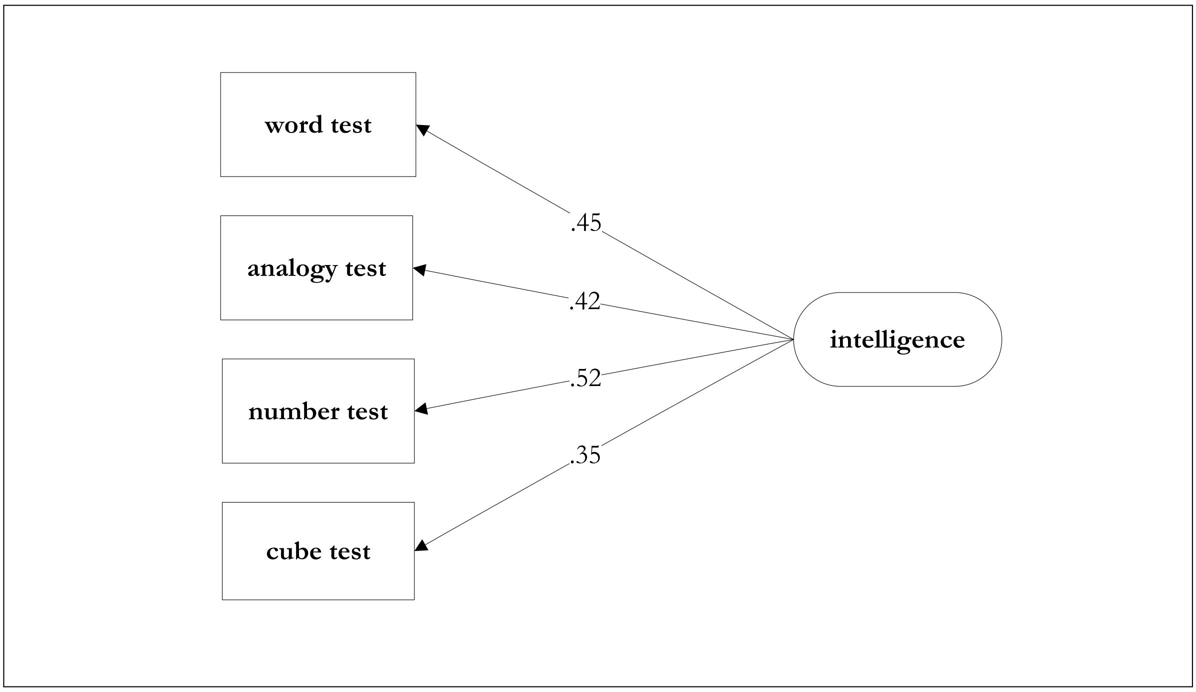


Figure 3: IST measurement model.

[26] All regression weights and covariances that are displayed in this and the subsequent structural equation figures (figures 3-6) show corresponding z-values that fulfill a significance value of $p < .05$ or lower (two-tailed).

[27] The following cut-off values for the goodness-of-fit criteria are convenient (Hu and Bentler, 1999; but also see Chen et al., 2008): $AGFI > .95$, $CFI > .90$, and both $RMSEA$ and $SRMR < .08$ (better $< .05$).

4.2 Structural Part

In the structural part, we will proceed in three subsequent steps that mainly follow the order of our hypotheses in section 2.2: First, teachers' evaluations are regressed on students' average grade. We label this model *performance model 1*. Second, this single-arrow model is amended by the latent intelligence variable as it has been estimated in the IST measurement model. This model is labeled *performance model 2*. Third, the SES indicators are introduced in order to model the *primary* effects of social inequality (*SES model*). And finally, also students' aspirations are included also to model the *secondary* effects of social inequality (*aspiration model*). According to our theoretical considerations, the indicators for both models may take effect via both modes of information processing.

Performance models The *performance model 1* simply regresses teachers' evaluations (1 = 'not able'; 2 = 'not mentioned'; 3 = 'able') on students' average grade. The standardized covariance of these two variables is about .30; and since the model is saturated, the fit measures of $PERF_1$ are perfect (table 1). Due to the simplicity of the model, we see no need for graphical illustration.

Performance model 2 extends performance model 1 in adding students' latent intelligence variable as a second exogenous variable (table 1, model $PERF_{2a}$). In our theoretical section, we expected that we might find stronger or additional effects for the verbal part of our intelligence test. And indeed, modification indices (Sörbom, 1989) suggested to allow for a direct covariance between the IST sub-score measuring the analogy-based dimension of intelligence and teachers' evaluations. Since it does not make much sense to assume a *cross-sectional* impact of teacher evaluations on students' intelligence[28], we only allowed for a one-way relationship in terms of an impact of intelligence on teachers' evaluations.[29] Table 1 indicates that this step clearly improves the fit of our model.

Moreover, first we did not allow for a covariance between intelligence and average grade – although, according to our theoretical considerations, we surely expected it to exist. The fit of the constrained model $PERF_{2b}$ was not very satisfactory, and thus we followed our theoretical assumptions and allowed for a one-way coefficient of intelligence on average grade. The fit of this model was a bit better (model $PERF_{2c}$), but could still be improved: Interestingly, modification indices also suggested another *direct* effect of the analogy sub-score on students' average grade (which seems to confirm our hypothesis

[28]In contrast, several studies modeled the Pygmalion effect as a *longitudinal* impact of teachers' evaluations on intelligence (e.g. Rosenthal and Jacobson, 1968 and all studies analyzed in Smith, 1980). Others focused on changes in school grades while controlling for intelligence (e.g. Smith et al., 1999). Although we are not directly testing the self-fulfilling prophecy hypothesis, we will yet consider its basic idea in terms of a covariance between teacher evaluations and school grades.

[29]Jöreskog (1993, p. 312) strongly recommends only to relax parameters which can be interpreted substantively. In our case, two arguments seem to make sense. Possibly, the competence of a student to draw analogy-based inferences is more applicable (and thus also more visible to teachers) in school lessons than the other sub-dimensions of intelligence. Another explanation would be that teachers rate students' competence in drawing analogy-based inferences particularly high with respect to successfully completing academic studies.

about the particular visibility of this sub-dimension of intelligence at school). This model, $PERF_{2d}$, is presented in figure 4. The numbers next to the arrows show the

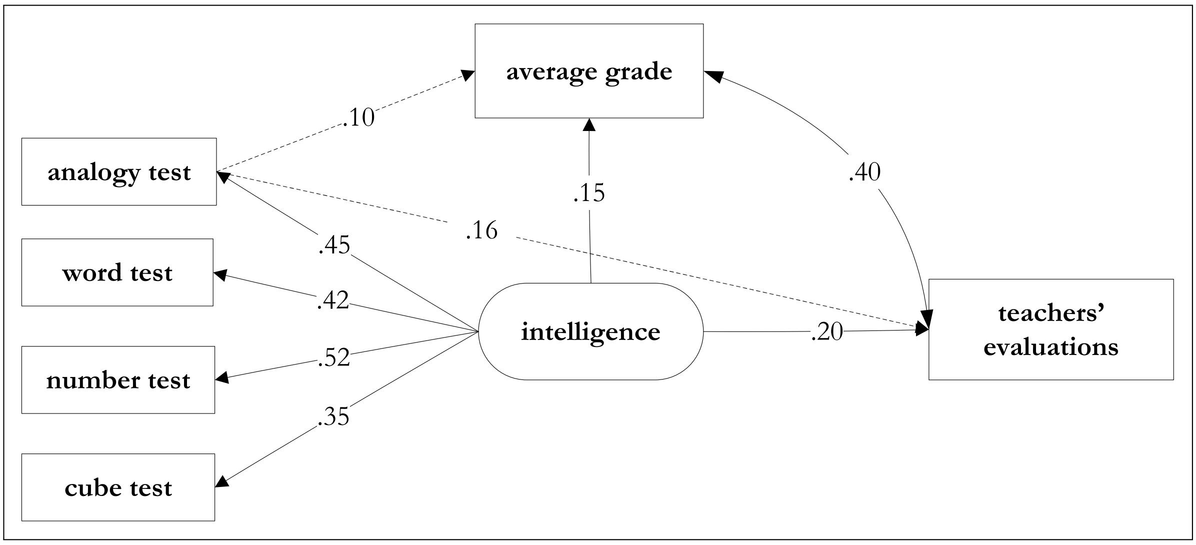


Figure 4: Performance model.

standardized path coefficients, the factor loadings, and the covariances of the model, respectively. Similar to our logistic regressions (cf. appendix, tables B and C), the covariance between average grade and teachers' evaluations seems to be much larger than the impact of students' intelligence scores (.40 vs. .20). Controlling also for intelligence, we note that the relationship between intelligence and teachers' evaluations is mediated by the intervening variable average grade (.15). It also seems noteworthy that the distinct effect of the IST sub-score measuring the analogy dimension of intelligence on average grade (.10) is again not much smaller than the respective overall regression weight of the latent variable intelligence (.15) – which is due to a drop-down of the latter from .43 in the restricted model (not shown). The fit of this model is convincing (cf. table 1, model $PERF_{2d}$).

SES model Now we introduce the maximum value of both highest parental educational degree and occupational prestige as two single indicators in order to model the *primary effects* of social inequality explicitly. The initial fit of this model is already acceptable (see table 2, model SES_1), and it could be improved slightly when the covariance between the two SES indicators was relaxed (model SES_2). Another improvement could be achieved when we allowed for the regression weights of the two SES indicators on the latent intelligence variable (model SES_3) – meaning an operationalization of *primary* effects of social inequality. Yet, in contrast to our theoretical model (figure 2), two coefficients in the SES model turned out to lack statistical significance: the coefficient of education on the overall latent intelligence variable, and the coefficient of occupational prestige on teachers' evaluations. Therefore, we subsequently dropped these regression weights (models SES_4 and SES_5). Moreover, modification indices suggested to intro-

Table 1: Performance models: Fit measures

	$PERF_1$	$PERF_{2a}$	$PERF_{2b}$	$PERF_{2c}$	$PERF_{2d}$
χ^2	<.001	172.03	138.02	28.015	8.85
df	1	10	9	7	6
$p(>\chi^2)$	1	<.001	<.001	<.001	0.182
$AGFI$	1	0.966	0.97	0.992	0.997
$RMSEA$	0	0.069	0.065	0.03	0.012
CFI	1	0.912	0.93	0.989	0.998
$SRMR$	0	0.061	0.058	0.018	0.009

duce a direct effect of parental education on the analogy sub-score of intelligence. Since we already found direct effects of this dimension on both average grade and teachers' evaluations (see figure 4), which was in line with our theoretical considerations, we allowed for this regression weight (model SES_6). While models SES_5 and SES_6 still contain occupational prestige as a covariate of education, we finally tested a model that completely passed the former variable (model SES_7). This model could achieve a better fit than SES_6, and, according to Occam's razor's maxim of parsimoniousness, it is the preferred model up to now (see figure 5).[30] The direct effect of parental education on teachers' evaluations is about .10 – which is the second smallest coefficient in the model up to now. Yet, we also have to keep in mind the indirect effect in terms of the relationship between education, the IST analogy dimension and teachers' evaluations.[31] The covariance between students' average grade and teachers' evaluations is still the strongest effect in the model (.40), while – at least up to now – the impact of intelligence on teachers' evaluations comes second (.26). Again, the effect of the latent intelligence variable on teachers' evaluations slightly increases when controlling for direct and indirect effects of parental education. Apparently, the predictive power of intelligence on teachers' evaluations becomes even stronger among students with the same social background. This model yielded an entirely satisfactory fit (table 2, model *SES7*).

Aspiration model In order to model also the *secondary effects* of social inequality as well as teachers' more or less conscious reflections on them, we finally included students' aspirations measured by their dummy-coded appraisement if 'Abitur' is necessary to

[30] We tested three additional variants of models SES_6 and SES_7 (not shown, available upon request): one with a regression weight of occupational prestige on average grade (not significant), one with a direct effect of education on the latent IST variable rather than on the sub-score merely measuring its analogy dimension (significant, but worse model fit), and one with regression weights of parental education on both the latent IST variable and the analogy sub-score (which is significant, but suffers from multicollinearity). Because of these respective drawbacks we still prefer model SES_7.

[31] The total effect of parental education on teachers' evaluations is computed as $.10 + (.11 \cdot .08 \cdot .4) + (.11 \cdot .13) \approx .118$.

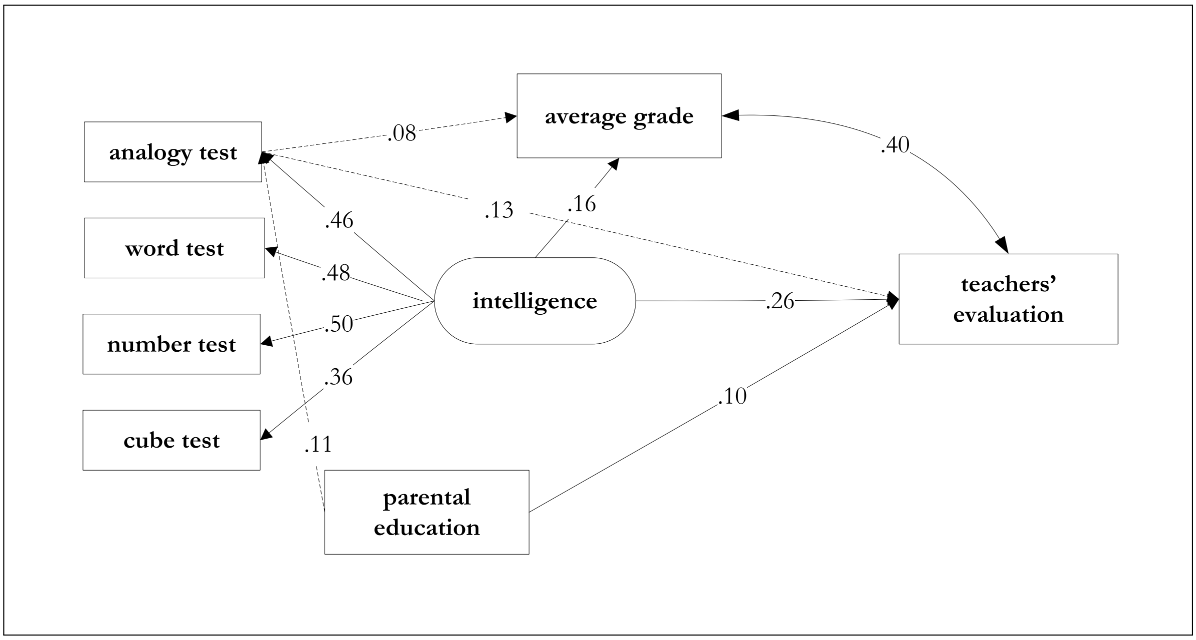


Figure 5: SES model.

Table 2: SES models: Fit measures

	SES_1	SES_2	SES_3	SES_4	SES_5	SES_6	SES_7
χ^2	1563.100	78.170	69.152	69.394	75.486	38.477	22.746
df	19	18	16	17	18	17	11
$p(>\chi^2)$	$<.001$	<0.001	<0.001	<0.001	<0.001	0.002	0.019
$AGFI$	0.835	0.989	0.989	0.989	0.989	0.994	0.995
$RMSEA$	0.155	0.031	0.031	0.030	0.031	0.019	0.018
CFI	0.556	0.983	0.985	0.985	0.983	0.994	0.994
$SRMR$	0.103	0.026	0.022	0.022	0.023	0.015	0.013

reach their aim in life. The fit of the initial model without allowing any additional covariances or regression weights except the direct effect of students' aspirations on teachers' evaluations (table 3, model ASP_1) could be improved when we allowed for a regression weight of education on students' aspirations (model ASP_2). Furthermore, we postulated a direct effect of intelligence on aspirations – which once more upgraded the fit of our model (model ASP_3). Next to these additional arrows, we also hypothesized a covariance between students' aspirations and their average grade. However, in the model including this covariance, it turned out to lack statistical significance (not shown, available on request). Therefore, ASP_3 is already our final model (figure 6).

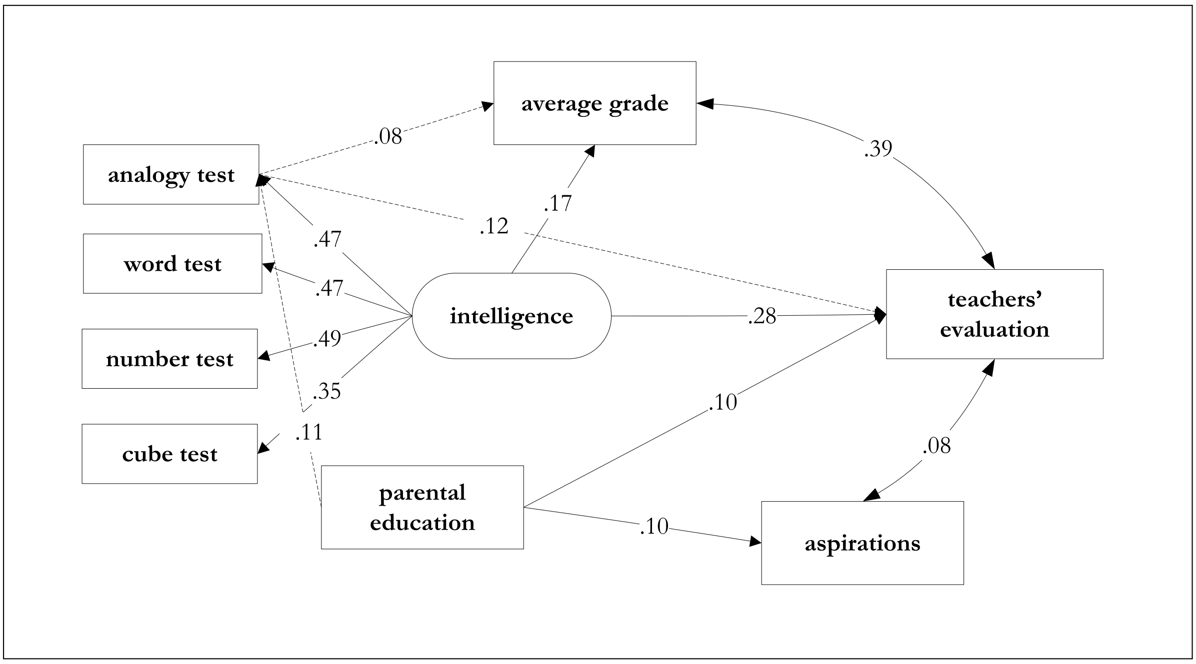


Figure 6: Aspiration model.

The largest effect in our model is still the covariance between average grade and teachers' evaluations (.39) while the regression weight of the latent intelligence variable comes second (.28). The covariance between students' aspirations and teachers' evaluations, however, is far lower (.08). Aspirations, in turn, are significantly predicted by parental education (.10) and students' intelligence (.14). Given the size of the final model, its fit is very satisfactory (table 3, model *ASP3*).

5 Summary and Outlook

In this paper, we analyzed the emergence of 10^{th} class teachers' evaluations with regard to students' prospective aptitude for academic studies.

In the theoretical section, we first tried to locate teachers' evaluations within the underlying social situation in the classroom (Coleman, 1990; Esser, 1993, 1996, 1999), and we then synthesized available meritocratic and *habitus*-related explanations of teachers'

Table 3: Aspiration models: Fit measures

	ASP_1	ASP_2	ASP_3
χ^2	102.11	70.197	33.417
df	28	28	28
$p(> \chi^2)$	$<$.001	$<$.001	.007
$AGFI$	.985	.989	.995
$RMSEA$	.037	.030	.018
CFI	.959	.974	.992
$SRMR$	.033	.028	.016

evaluations to a more general theory of action in terms of the *M*odel of *F*rame *S*election (MFS; Esser, 1996, 2010; Esser and Kroneberg, 2010; Kahneman and Tversky, 1984; Kroneberg, 2006; Kroneberg et al., 2008, 2010; Kroneberg, 2011; Kroneberg and Kalter, 2012).

Subsequent to that endeavor, in a short literature review we derived four hypotheses postulating that teachers' evaluations would be influenced by students' intelligence, average grade, social backgrounds, and aspirations, respectively. Furthermore, we expected that some of these independent variables would show a path structure in terms of additional regression weights or covariances between them (Figure 2).

This model was tested by means of the "*C*ologne *Hi*gh *S*chool *P*anel" (*CHiSP*, 1969). From logistic regression analyses (tables B and C), we could already observe that students' average grade has the strongest impact on (positive or negative) teachers' evaluations – while their aspirations come second. Another result logistic regression pointed to was that the category of receiving no evaluation at all apparently is located between the category of obtaining a positive evaluation and the one of getting a negative one. Therefore, for the subsequent structural equation models as our main analyses, we modeled teachers' evaluations of their students' ability for academic studies as our dependent variable in the following way: 1 'not able'; 2 'not mentioned'; 3 'able'.

In the structural equation models our main hypotheses were corroborated. Even when controlling for the expected path structure, all of our (formally) independent variables showed significant effects on teachers' evaluations. Average grade is still the strongest predictor, but in contrast to preceding logistic regression analyses, now students' intelligence comes second, and their aspirations come third.

In addition to our main hypothesis H_2, we already found evidence in the literature that the verbal dimension of intelligence might be more important for teachers' evaluations than the numeric dimension. Indeed, we could observe independent effects of the analogy sub-score of intelligence on both average grade and teachers' evaluations. If one would try to link that finding back to our theoretical considerations rooted in the MFS, this dimension could at least partially reflect either a meritocracy-based *rc-mode* or even a *habitus*-based *as-mode* of processing.[32]

But compared to the initial path model, we also had to drop several arrows due to lack of significance: First, we did not find a significant regression weight of parental education on the overall latent intelligence variable. However, we could note a significant impact of parental education on the analogy sub-score of intelligence. Since this variable showed independent effects on both average grade and teachers' evaluations, we conclude that the *primary* effect of social inequality is mainly passed on via this predictor. Second, we could not find any direct effects of parental occupational prestige on students' average grade. Apparently, in our socially selective sample – recall that our observations are (predominantly upper-class) *Gymnasium* students – the primary effect of social inequality is exhaustively modeled when we control for the indirect effect of parental SES via intelligence on average grade. The third arrow we had to drop concerned the regression weight of parental occupational prestige on students' aspirations. This indicates that by controlling for parental education, all social background effects on students' aspirations are already modeled.

In sum, we can conclude that although indicators for all four types of theoretical concepts showed statistical significance, results confirmed our expectations that the meritocracy explanation – be it based on *rc-mode* or *as-mode* scripts – shows more predictive power than the explanation based on *habitus*-related criteria. Yet, both the empirical dominance of students' average grade in our models and the fact that the verbal dimension of intelligence showed additional path coefficients on both average grade and teachers' evaluations might underline the particular importance of the *as-mode* type of meritocratic reasoning.

These results suggest the following implications for further studies: First, the underlying social mechanisms of the emergence of teachers' evaluations have to be further examined. Future studies should try to sharpen the distinction between *rc-mode* and *as-mode* processing type explanations as we have transferred it on the social situation in the classroom.

Clearly, this approach needs the consideration of more background variables. On the one hand, the set of student variables in our analyses might be no exhaustive operationalization of the student side of the social situation in the classroom. Thus, it would make sense to include additional information such as students' grades in different subjects or their academic self-concept in order to specify the social situation in the classroom more concretely.

Moreover, we already indicated that although at first sight, it appears reasonable to interpret the additional path coefficient of the analogy sub-score on both students' academic performance and teachers' evaluations in line with an automatically-driven meritocratic form of reasoning, at a second glance, these arrows might also emerge by virtue of unconscious, symbolic communication among teachers and students usually referred to as effects of *habitus*: Recapitulating our theoretical considerations strictly in the latent variable framework, only one of our lower-level concepts intelligence, academic

[32]Below we discuss why we see arguments for either mode of processing, and we propose a method how to decide which mode of processing may be the actual drive for this arrow.

performance, parental SES and students aspirations was actually measured as a latent variable, namely students' intelligence. Thus, both students' academic performance and their social backgrounds – more or less consciously observed by their teachers – were operationalized by single manifest indicators that probably did not control sufficiently for measurement error. In other words, students' *objective* academic performance should be understood as a latent variable which is only approximately measured by their average grades. The latter, in turn, are nothing but the result of a specific form of teachers' evaluations which may *themselves* be distorted by unconscious, *habitus*-related criteria that operate in addition to teachers' meritocratic considerations. We expect that we probably *would* find supplemental effects of both the verbal dimension of intelligence and our measure of academic performance on our indicators of students' social backgrounds – currently only in part reflecting teachers' unconscious mental processes – if we could provide a more detailed operationalization of the efficacy of *habitus* (e.g. in terms of students' cultural capital, their cultural practices, etc.) than we were able to measure with our data at hand.[33] In concrete terms, we demand from further studies to test for a second-order factor model (Chen et al., 2005; Rindskopf and Rose, 1988) with students' *habitus* as the higher-level factor, and parental SES, students' aspirations and their cultural capital as lower-level factors that should be operationalized by appropriate indicators, respectively.[34]

On the other hand, if one would really want to disentangle the conditions determining in which situations teachers' scripts of action tend to follow either a more automatic or a more rational mode of information processing – which we solely sketched for theoretical elucidation purposes –, it will be inevitable also to control for teacher background variables. Future studies should try to find variables such as teachers' pedagogic concepts, their attitudes towards educational inequality or measures of teachers' success attribution that explain why a particular teacher follows a certain dominant script of action. Furthermore, teachers' backgrounds should ideally cover indicators of the symbolic processes usually referred to as effects of *habitus* as well: Only if both students' and teachers' *habitus* are measured adequately, a final decision about *habitus* match or mismatch will be possible. Methodologically, controlling also for teachers' backgrounds

[33] See Kingston (2001), Lareau and Weininger (2003), and, recently, Andersen and Hansen (2012) for a critical assessment of cultural capital usage in educational research.

[34] DiMaggio (1982) and De Graaf (1986) used exploratory factor analysis to measure families' cultural capital, but to the best of our knowledge, a confirmatory factor model of the notion of *habitus* in a broader sense is still missing. McClelland (1990), Dumais (2002, 2006) and Andres (2009) operationalized *habitus* by students' aspirations, but as Dumais (2002, p. 51) herself acknowledged, single-indicator measures for *habitus* are far from perfect (also see Reay, 2004, p. 440f.). Andres (2009) made use of a path model to test the interrelations between social backgrounds, different forms of capital and dispositions; but although claimed in his theoretical section, no analytical operationalization of *habitus* was given in his measurement part. In this attempt, further studies may also refer to the theoretical concepts used by social psychology that offer a whole bunch of literature about the prediction of behavior by attitudes (for meta-analyses see Glasman and Albarracín, 2006; Kim and Hunter, 1993a,b; Kraus, 1995; Wallace et al., 2005). However, although Acock and Scott (1980) already modeled attitudes as being affected by social class, more recent psychological studies apparently neglected this endogeneity.

would equal a multilevel structural equation model (Bauer, 2003; Heck, 2001; Muthén, 1994; Rabe-Hesketh et al., 2004, 2007) wherein both students and teachers' evaluations are nested in teacher contexts.

And finally, the theoretical and methodological propositions offered here would certainly gain from a more comparative framework.[35] Provided that reliable indicators allowing to disentangle both modes of processing from each other can be found, it would be illuminating to see whether the potential prevalence of either one is contingent upon differences in social selectivity *between* educational systems, or upon changes in social selectivity and/or saturation *within* a given one (for an overview see Shavit and Blossfeld, 1993).

References

Acock AC and Scott WJ (1980) A model for predicting behavior: the effect of attitude and social class on high and low visibility political participation. *Social Psychology Quarterly* 43(1): 59–72.

Aish AM and Jöreskog KG (1990) A panel model for political efficacy and responsiveness: an application of LISREL 7 with weighted least squares. *Quality and Quantity* 24(4): 405–426.

Alexander KL; Entwisle DR and Thompson MS (1987) School performance, status relations, and the structure of sentiment: bringing the teacher back in. *American Sociological Review* 52(5): 665–682.

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

Andersen PL and Hansen MN (2012) Class and cultural capital - the case of class inequality in educational performance. *European Sociological Review* 28(5): 607–621.

Andres L (2009) The cumulative impact of capital on dispositions across time: a 15 years perspective of young canadians. In: *Quantifying Theory: Pierre Bourdieu*, 75–88. Dordrecht: Springer Netherlands.

Arnold KH; Bos W; Richert P et al. (2007) Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In: Bos W; Hornberg S; Arnold KH et al. (eds.) *IGLU 2006 - Schullaufbahnempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe*, 271–297. Münster / New York: Waxmann.

Babakus E; Ferguson Jr CE and Jöreskog KG (1987) The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research* 24(2): 222–228.

[35] This would be of particular relevance also regarding the MFS since in a comparative framework, variance in teachers' frames of the underlying social situation could be studied as well.

Baeriswyl F; Wandeler C; Trautwein U et al. (2006) Leistungstest, Offenheit von Bildungsgängen und obligatorische Beratung der Eltern. *Zeitschrift für Erziehungswissenschaft* 9(3): 373–392.

Bauer DJ (2003) Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics* 28(2): 135–167.

Becker D (2010) The impact of teachers' expectations on students' educational opportunities in the life course. *Cologne Graduate School Working Paper Series* 1(7). **URL:** *http://econpapers.repec.org/paper/cgrcgsser/01-07.htm*

Becker D and Birkelbach K (2010) Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond- oder Reflected-Glory-Effekt? Eine Mehrebenen-Analyse von Lehrerurteilen. In: Beckers T; Birkelbach K; Hagenah J et al. (eds.) *Komparative empirische Sozialforschung*, 113–141. Wiesbaden: VS-Verlag.

Becker R (2003) Educational expansion and persistent inequalities of education. *European Sociological Review* 19(1): 1–24.

Becker R and Hadjar A (2011) Meritokratie - Zur gesellschaftlichen Legitimation ungleicher Bildungs-, Erwerbs-und Einkommenschancen in modernen Gesellschaften. *Lehrbuch der Bildungssoziologie* 37–62.

Becker R and Hecken AE (2009) Why are working-class children diverted from universities? An empirical assessment of the diversion thesis. *European Sociological Review* 25(2): 233–250.

Bell D (1972) On Meritocracy and Equality. *The Public Interest* 29: 29–68.

Birkelbach K (1998) *Berufserfolg und Familiengründung. Lebensläufe zwischen institutionellen Bedingungen und individueller Konstruktion.* Opladen: Westdeutscher Verlag.

Birkelbach K (2011) Teacher evaluations over the life course: valid prognosis or self-fulfilling prophecy? In: Mica A; Peisert A and Winczorek J (eds.) *Sociology and the Unintended - Robert Merton Revisited*, 133–153. Frankfurt a.M.: Peter Lang.

Bobertag O and Hylla E (1926) *Begabungsprüfung für den Übergang von der Grundschule zu weiterführenden Schulen.* Langensalza: Verlag von Julius Beltz.

Bollen KA (1989) *Structural Equations With Latent Variables.* New York: Wiley.

Bollen KA and Lennox R (1991) Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin* 110(2): 305–314.

Bos W and Pietsch M (2004) *Erste Ergebnisse aus KESS 4. Kurzbericht.* Hamburg: Freie und Hansestadt Hamburg.

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

Bourdieu P (1973) Cultural reproduction and social production. In: Brown R (ed.) *Knowledge, Education, and Cultural Change: Papers in the Sociology of Education*, 71–112. London: Tavistock Publications.

Bourdieu P (1986) *Distinction: A Social Critique of the Judgement of Taste.* London: Routledge.

Bourdieu P and Passeron JC (1990) *Reproduction in Education, Society and Culture.* London: Sage Publications Ltd.

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 Goldthorpe JH (1999) Class inequality and meritocracy: a critique of Saunders and an alternative analysis. *The British Journal of Sociology* 50(1): 1–27.

Breen R and Jonsson JO (2000) Analyzing educational careers: a multinomial transition model. *American Sociological Review* 65(5): 754–772.

Breen R; Luijkx R; Müller W et al. (2009) Nonpersistent inequality in educational attainment: evidence from eight European countries. *American Journal of Sociology* 114(5): 1475–1521.

Breen R; Luijkx R; Muller W et al. (2010) Long-term trends in educational inequality in Europe: class inequalities and gender differences. *European Sociological Review* 26(1): 31–48.

Burton NW and Ramist L (2001) *Predicting Success in College: SAT Studies of Classes Graduating Since 1980.* New York: College Board Research Report No. 2001-2.

Byrne BM (2004) Testing for multigroup invariance using AMOS graphics: a road less traveled. *Structural Equation Modeling: A Multidisciplinary Journal* 11(2): 272–300.

Camara W (1998) *High School Grading Policies.* Tech. Rep. RN-04, College Board, New York.

Camara W and Echternacht G (2000) *The SAT I and High School Grades: Utility in Predicting Success in College.* New York: The College Board.

Camara W; Kimmel E; Scheuneman J et al. (2003) *Whose Grades Are Inflated?* Tech. Rep. RN-10, College Board, New York.

Chen F; Curran PJ; Bollen KA et al. (2008) An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research* 36(4): 462–494.

Chen FF; Sousa KH and West SG (2005) Teacher's corner: testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal* 12(3): 471–492.

Coleman JS (1990) *Foundations of Social Theory.* New York: Harvard University Press.

Collet F (2009) Does habitus matter? A comparative review of Bourdieu's habitus and Simon's bounded rationality with some implications for economic sociology. *Sociological Theory* 27(4): 419–434.

De Graaf ND and De Graaf PM (2002) Formal and popular dimensions of cultural capital: effects on children's educational attainment. *The Netherlands' Journal of Social Sciences* 38(2): 167–186.

De Graaf ND; De Graaf PM and Kraaykamp G (2000) Parental cultural capital and educational attainment in the Netherlands: a refinement of the cultural capital perspective. *Sociology of Education* 73(2): 92–111.

De Graaf PM (1986) The impact of financial and cultural resources on educational attainment in the Netherlands. *Sociology of Education* 59(4): 237–246.

Diamantopoulos A and Winklhofer HM (2001) Index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research* 38(2): 269–277.

DiMaggio P (1982) Cultural capital and school success: the impact of status culture participation on the grades of U.S. high school students. *American Sociological Review* 47(2): 189–201.

Ditton H (2007) Der Beitrag von Schule und Lehrern zur Reproduktion von Bildungsungleichheit. In: Becker R and Lauterbach W (eds.) *Bildung als Privileg*, 247–275. Wiesbaden: VS Verlag für Sozialwissenschaften.

Downey DB and Pribesh S (2004) When race matters: teachers' evaluations of students' classroom behavior. *Sociology of Education* 77(4): 267–282.

Dronkers J (1983) Have inequalities in educational opportunity changed in the Netherlands? A review of empirical evidence. *The Netherlands Journal of Sociology* 19(2): 133–150.

Dronkers J (1993) Educational reform in the Netherlands: did it change the impact of parental occupation and education? *Sociology of Education* 66(4): 262–277.

Dumais SA (2002) Cultural capital, gender, and school success: the role of habitus. *Sociology of Education* 75(1): 44–68.

Dumais SA (2006) Early childhood cultural capital, parental habitus, and teachers' perceptions. *Poetics* 34(2): 83–107.

Elster J (1983) *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press.

Erikson RS (1984) Social class of men, women and families. *Sociology* 18(4): 500–514.

Erikson RS (2007) Social selection in Stockholm schools: primary and secondary effects on the transition to upper secondary education. In: Scherer S; Pollak R; Otte G et al. (eds.) *From Origin to Destination: Trends and Mechanisms in Social Stratification Research*, 61–81. Frankfurt a.M.: Campus.

Erikson RS; Goldthorpe JH; Jackson M et al. (2005) On class differentials in educational attainment. *Proceedings of the National Academy of Sciences of the United States of America* 102(27): 9730–9733.

Erikson RS and Jonsson JO (1996) Introduction: explaining class inequality in education: the Swedish test case. In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 1–63. Boulder, CO: Westview Press.

Esser H (1993) *Soziologie. Allgemeine Grundlagen.* Frankfurt / New York: Campus Verlag.

Esser H (1996) Die Definition der Situation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 48(1): 1–34.

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

Esser H (2010) Das Modell der Frame-Selektion. Eine allgemeine Theorie der Sozialwissenschaften. In: Albert G and Sigmund S (eds.) *Soziologische Theorie kontrovers (Sonderheft 50 der Kölner Zeitschrift für Soziologie und Sozialpsychologie)*, 50–62. Wiesbaden: VS Verlag für Sozialwissenschaften.

Esser H and Kroneberg C (2010) Am besten nichts Neues? In: Albert G and Sigmund S (eds.) *Soziologische Theorie kontrovers (Sonderheft 50 der Kölner Zeitschrift für Soziologie und Sozialpsychologie)*, 79–86. Wiesbaden: VS Verlag für Sozialwissenschaften.

Farkas G (2003) Racial disparities and discrimination in education: what do we know, how do we know it, and what do we need to know? *The Teachers College Record* 105(6): 1119–1146.

Farkas G; Grobe RP; Sheehan D et al. (1990) Cultural resources and school success: gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review* 55(1): 127–142.

Ferguson RF (2003) Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education* 38(4): 460–507.

Fox J (2006) Structural equation modeling with the sem package in R. *Structural Equation Modeling: A Multidisciplinary Journal* 13(3): 465–486.

Ganzeboom HBG and Treiman DJ (1996) Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25(3): 201–239.

Glaesser J and Cooper B (2011) Selectivity and flexibility in the German secondary school system: a configurational analysis of recent data from the German Socio-Economic Panel. *European Sociological Review* 27(5): 570–585.

Glasman LR and Albarracín D (2006) Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychological bulletin* 132(5): 778–822.

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 (2003) The myth of education-based meritocracy. *New Economy* 10(4): 234–239.

Goldthorpe JH (2007) "Cultural Capital": Some Critical Observations. *Sociologica* 1(2): 1–23.

Goldthorpe JH and Jackson M (2008) 'Education-based meritocracy: the barriers to its realization. In: Lareau A and Conley D (eds.) *Social Class: How Does it Work?*, 93–117. New York: Russell Sage Foundation Publication.

Gröhlich C and Guill K (2009) How stable are reference group effects of secondary school track recommendations? *Journal for Educational Research Online/Journal für Bildungsforschung Online* 1(1): 154–171.

Heck RH (2001) Multilevel modeling with SEM. In: Marcoulides GA and Schumacker RE (eds.) *New Developments and Techniques in Structural Equation Modeling*, 89–127. Hillsdale: Lawrence Erlbaum.

Hedström P (2005) *Dissecting The Social. On the Principles of Analytical Sociology.* Cambridge: Cambridge University Press.

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.

Hoge RD and Coladarci T (1989) Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research* 59(3): 297–313.

Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6(1): 1–55.

Hughes JN; Gleason KA and Zhang D (2005) Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology* 43(4): 303–320.

Ingenkamp K (1971) *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte.* Weinheim & Basel: Beltz Verlag.

Jaeger MM (2009) Equal access but unequal outcomes: cultural capital and educational choice in a meritocratic society. *Social Forces* 87(4): 1943–1971.

Jöreskog KG (1993) Testing structural equation models. In: Bollen KA and Long JS (eds.) *Testing Structural Equation Models*, 294–316. London: Sage Publications Ltd.

Jöreskog KG (1994) On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* 59(3): 381–389.

Jöreskog KG and Sörbom D (1989) *LISREL 7. A guide to the program and applications.* Chicago, Ill.: SPSS Inc.

Jürges H and Schneider K (2006) Age at school entry and teachers' recommendations for secondary school track choice in Germany. *Paper presented at the 20th annual ESPE conference, 22 - 24 June, 2006, Verona* .

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.

Kahneman D and Tversky A (1984) Choices, values, and frames. *American Psychologist* 39(4): 341–350.

Kaplan LD and Kaplan C (1997) Democracy, meritocracy, and the cognitive elite: the real thesis of the bell curve. *Educational Theory* 47(3): 425–431.

Kelle U and Lüdemann C (1995) Grau, teurer Freund, ist alle Theorie... Rational Choice und das Problem der Brückenannahmen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 47(2): 249–267.

Kim MS and Hunter JE (1993a) Attitude-behavior relations: a meta-analysis of attitudinal relevance and topic. *Journal of Communication* 43(1): 101–142.

Kim MS and Hunter JE (1993b) Relationships among attitudes, behavioral intentions, and behavior. *Communication Research* 20(3): 331–364.

Kingston PW (2001) The unfulfilled promise of cultural capital theory. *Sociology of Education* 74: 88–99.

Kraus SJ (1995) Attitudes and the prediction of behavior: a meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin* 21(1): 58–75.

Kroneberg C (2005) Die Definition der Situation und die variable Rationalität der Akteure: Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie* 34: 344–363.

Kroneberg C (2006) The definition of the situation and variable rationality: The model of frame selection as a general theory of action. *Sonderforschungsbereich 504 Publications, University of Mannheim* .

Kroneberg C (2011) *Die Erklärung sozialen Handelns. Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kroneberg C; Heintze I and Mehlkop G (2008) On shoplifting and tax fraud: An action-theoretic analysis of crime. *Sonderforschungsbereich 504 Publications, University of Mannheim* .

Kroneberg C and Kalter F (2012) Rational Choice Theory and Empirical Research: Methodological and Theoretical Contributions in Europe. *Annual Review of Sociology* 38: 73–92.

Kroneberg C; Yaish M and Stocké V (2010) Norms and rationality in electoral participation and in the rescue of Jews in WWII. *Rationality and Society* 22(1): 3–36.

Kuncel NR; Hezlett SA and Ones DS (2001) A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: implications for graduate student selection and performance. *Psychological Bulletin* 127(1): 162–181.

Lareau A and Weininger EB (2003) Cultural capital in educational research: a critical assessment. *Theory and Society* 32(5): 567–606.

Lörz M and Schindler S (2009) Educational expansion and effects on the transition to higher education: Has the effect of social background characteristics declined or just moved to the next stage? In: Hadjar A and Becker R (eds.) *Expected and Unexpected Consequences of the Educational Expansion in Europe and USA*, 97–110. Bern: Haupt.

Lucas SR (2001) Effectively maintained Iinequality: education transitions, track mobility, and social background effects. *American Journal of Sociology* 106(6): 1642–1690.

Lucas SR (2009) Stratification theory, socioeconomic background, and educational attainment. *Rationality and Society* 21(4): 459–511.

MacCallum RC and Browne MW (1993) The use of causal indicators in covariance structure models: some practical issues. *Psychological Bulletin* 114(3): 533–541.

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.

Mayer KU; Müller W and Pollak R (2007) Germany: Institutional change and inequalities of access in higher education. In: Shavit Y; Arum R and Gamoran A (eds.) *Stratification in Higher Education. A Comparative Study*, 240–265. Stanford: Stanford University Press.

McClelland K (1990) Cumulative disadvantage among the highly ambitious. *Sociology of Education* 63(2): 102–121.

Meulemann H; Birkelbach K and Hellwig O (2001) *Ankunft im Erwachsenenleben. Lebenserfolg und Erfolgsdeutung in einer Kohorte ehemaliger Gymnasiasten zwischen 16 und 43.* Opladen: Leske + Budrich Verlag.

Milek A; Lüdtke O; Trautwein U et al. (2009) Wie konsistent sind Referenzgruppeneffekte bei der Vergabe von Schulformempfehlungen? Bundeslandspezifische Analysen mit Daten der IGLU-Studie. *Zeitschrift für Erziehungswissenschaft* 12(12): 282–301.

Moede W; Piorkowski C and Wolff G (1919) *Die Berliner Begabtenschulen, ihre Organisation und die Methoden ihrer Schülerauswahl.* Langensalza: Bayer.

Morgan R (1989) *Analyses of the Predictive Validity of the SAT and High School Grades From 1976 to 1985.* New York: College Board Research Report No. 89-7.

Morgan SL (2002) Modeling preparatory commitment and non-repeatable decisions. *Rationality and Society* 14(4): 387–429.

Morris EW (2005) From "middle class" to "trailer trash:" teachers' perceptions of white students in a predominately minority school. *Sociology of Education* 78(2): 99–121.

Müller W and Karle W (1993) Social selection in educational systems in Europe. *European Sociological Review* 9(1): 1–23.

Müller W and Pollak R (2007) Weshalb gibt es so wenige Arbeiterkinder an deutschen Universitäten? In: Becker R and Lauterbach W (eds.) *Bildung als Privileg? Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit*, 303–342. Wiesbaden: VS Verlag für Sozialwissenschaften.

Muthén BO (1984) A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 49(1): 115–132.

Muthén BO (1994) Multilevel covariance structure analysis. *Sociological Methods & Research* 22(3): 376–398.

Olsson UH (1979) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 44(4): 443–460.

Page R (1987) Teachers' perceptions of students: a link between classrooms, school cultures, and the social order. *Anthropology & Education Quarterly* 18(2): 77–99.

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.

Rabe-Hesketh S; Skrondal A and Pickles A (2004) Generalized multilevel structural equation modeling. *Psychometrika* 69(2): 167–190.

Rabe-Hesketh S; Skrondal A and Zheng X (2007) Multilevel structural equation modeling. In: Lee SY (ed.) *Handbook of Latent Variable and Related models*, 209–227. Amsterdam: Elsevier.

Raftery AE and Hout M (1993) Maximally maintained inequality: expansion, reform, and opportunity in Irish education, 1921-75. *Sociology of Education* 66(1): 41–62.

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.

Ready DD and Wright DL (2011) Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. *American Educational Research Journal* 48(2): 335–360.

Reay D (2004) 'It's all becoming a habitus': beyond the habitual use of habitus in educational research. *British Journal of Sociology of Education* 25(4): 431–444.

Rigdon EE and Ferguson Jr CE (1991) The performance of the polychoric correlation coefficient and selected fitting functions in confirmatory factor analysis with ordinal data. *Journal of Marketing Research* 28(4): 491–497.

Rindskopf D and Rose T (1988) Second order factor analysis: some theory and applications. *Multivariate Behavioral Research* 23(1): 51–67.

Robbins SB; Lauver K; Le H et al. (2004) Do psychosocial and study skill factors predict college outcomes? a meta-analysis. *Psychological Bulletin* 130(2): 261–288.

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

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.

Schubert F and Becker R (2010) Social inequality of reading literacy - a longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pair wise matching procedure. *Research in Social Stratification and Mobility* 28(1): 109–133.

Selz M and Vallet LA (2006) La démocratisation de l'enseignement et son paradoxe apparent. *Données sociales - La société française* édition 20: 101–107.

Sewell WH; Haller AO and Ohlendorf GW (1970) The educational and early occupational status attainment process: replication and revision. *American Sociological Review* 35(6): 1014–1027.

Sewell WH; Haller AO and Portes A (1969) The educational and early occupational attainment process. *American Sociological Review* 34(1): 82–92.

Shavit Y and Blossfeld HP (1993) *Persistent Inequality. Changing Educational Attainment in Thirteen Countries*. San Francisco, Oxford: Westview Press.

Smith AE; Jussim L and Eccles J (1999) Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *Journal of Personality and Social Psychology* 77(3): 548–565.

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

Solga H (2005) Meritokratie - die moderne Legitimation ungleicher Bildungschancen. In: Berger PA and Kahlert H (eds.) *Institutionalisierte Ungleichheiten: Wie das Bildungswesen Chancen blockiert*, 19–38. Weinheim & München: Juventa.

Sörbom D (1989) Model modification. *Psychometrika* 54(3): 371–384.

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.

Tiedemann J and Billmann-Mahecha E (2007) Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 10(1): 108–120.

Tolsma J; Need A and de Jong U (2010) Explaining participation differentials in Dutch higher education: the impact of subjective success probabilities on level choice and field choice. *European Sociological Review* 26(2): 235–252.

Trapmann S; Hell B; Weigand S et al. (2007) Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie* 21(1): 11–27.

Trautwein U and Baeriswyl F (2007) Wenn leistungsstarke Klassenkameraden ein Nachteil sind. *Zeitschrift für Pädagogische Psychologie* 21(2): 119–133.

Treiman DJ (1977) *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

van de Werfhorst HG and Hofstede S (2007) Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *The British Journal of Sociology* 58(3): 391–415.

Wallace DS; Paulson RM; Lord CG et al. (2005) Which behaviors do attitudes predict? Meta-analyzing the effects of social pressure and perceived difficulty. *Review of General Psychology* 9(3): 214–227.

Weber M (1968) *Economy and Society: An Outline of Interpretive Sociology.* New York: Bedminster Press.

Yaish M and Katz-Gerro T (2012) Disentangling 'cultural capital': the consequences of cultural and economic resources for taste and participation. *European Sociological Review* 28(2): 169–185.

Zhang G and Browne MW (2006) Bootstrap fit testing, confidence intervals, and standard error estimation in the factor analysis of polychoric correlation matrices. *Behaviormetrika* 33(1): 61–74.

Ziegenspeck JW and Lehmann J (1999) *Handbuch Zensur und Zeugnis in der Schule.* Bad Heilbrunn: Klinkhart.

6 Appendix

Table A: Descriptive results

		valid	mean	stdev	min	max
teachers' evaluations		2427	2.06	0.75	1	3
	1 'not able'					
	2 'not mentioned'					
	3 'able'					
sex		3385	1.47	0.5	1	2
	1 'male'					
	2 'female'					
intelligence scores (overall index)		3230	110.45	11.35	76	151
analogy test		3230	111.66	11.66	77	152
analogy test (dichotomized)		3230	0.5	0.5	0	1
	0 'below median'					
	1 'above median'					
word test		3230	106.39	10.53	70	138
word test (dichotomized)		3230	0.48	0.5	0	1
	0 'below median'					
	1 'above median'					
number test		3230	106.82	10.93	80	147
number test (dichotomized)		3230	0.45	0.5	0	1
	0 'below median'					
	1 'above median'					
cube test		3230	103.21	10.76	73	140
cube test (dichotomized)		3230	0.47	0.5	0	1
	0 'below median'					
	1 'above median'					
average grade		3227	499.98	69.22	221	703
average grade (dichotomized)		3227	0.5	0.5	0	1
	0 'above median'					
	1 'below median'					
parental education (highest)		3374	2.14	1.23	1	4
	1 'lower'					
	2 'middle'					
	3 'Abitur'					
	4 'degree'					
occ. prestige (highest)		2687	49.37	12.63	18	78
occ. prestige (highest, dichotomized)		2687	0.47	0.5	0	1
	0 'below median'					
	1 'above median'					
aspirations		2543	2.43	0.71	1	3
	0 'no aim in life / Abitur not necessary'					
	1 'Abitur useful or necessary'					

Table B: Logistic regression: Able vs. not able

	Performance Model 1	Performance Model 2	SES Model	Aspiration Model
	e^b/z	e^b/z	e^b/z	e^b/z
constant	1.06	0.57*	0.30***	0.19***
	(0.29)	(-2.41)	(-4.08)	(-5.18)
sex	0.77*	0.53***	0.54***	0.58***
	(-2.21)	(-4.51)	(-3.79)	(-3.31)
intelligence	3.04***	2.33***	2.84***	2.79***
	(9.52)	(6.13)	(6.60)	(6.43)
average grade		12.35***	13.20***	12.93***
		(17.75)	(15.82)	(15.55)
parental education			1.19*	1.16
			(2.29)	(1.93)
parental occ. prestige			1.56*	1.59*
			(2.44)	(2.52)
aspirations				1.90***
				(3.78)
Nagelkerke's R^2	0.10	0.42	0.46	0.47
N	1314	1309	1067	1063

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

Table C: Logistic regression: Able vs. not mentioned

	Performance Model 1	Performance Model 2	SES Model	Aspiration Model
	e^b/z	e^b/z	e^b/z	e^b/z
constant	0.58***	0.28***	0.15***	0.12***
	(-3.30)	(-6.83)	(-8.15)	(-8.20)
sex	0.95	0.84	1.00	1.02
	(-0.56)	(-1.60)	(-0.02)	(0.16)
intelligence	1.77***	1.56***	1.79***	1.76***
	(5.71)	(4.17)	(4.93)	(4.73)
average grade		4.56***	4.72***	4.63***
		(13.41)	(12.27)	(12.10)
parental education			1.15*	1.13*
			(2.41)	(2.19)
parental occ. prestige			1.07	1.08
			(0.46)	(0.54)
aspirations				1.27
				(1.79)
Nagelkerke's R^2	0.03	0.17	0.19	0.19
N	1720	1716	1412	1406

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

Table D: Polychoric correlation matrix

	teachers' evaluations	*intelligence: word test*	*intelligence: analogy test*	*intelligence: number test*	*intelligence: cube test*	*average grade*	*max (education)*	*max (occ. prestige)*	*aspirations*
teachers' evaluations	1	0.15	0.26	0.17	0.1	0.48	0.12	0.11	0.16
intelligence: analogy test	0.15	1	0.24	0.23	0.15	0.1	-0.02	0.03	0.07
intelligence: word test	0.26	0.24	1	0.21	0.16	0.16	0.1	0.07	0.1
intelligence: number test	0.17	0.23	0.21	1	0.2	0.09	-0.04	0	0.06
intelligence: cube test	0.1	0.15	0.16	0.2	1	0.07	0.02	0.02	0.03
average grade	0.48	0.1	0.16	0.09	0.07	1	0.03	0.02	0.07
max (education)	0.12	-0.02	0.1	-0.04	0.02	0.03	1	0.6	0.1
max (occ. prestige)	0.11	0.03	0.07	0	0.02	0.02	0.6	1	0.09
aspirations	0.16	0.07	0.1	0.06	0.03	0.07	0.1	0.09	1

III. Intelligence and Academic Achievement as Contextual-Level Predictors of Teachers' Evaluations:

Big Fish Little Pond or Reflected Glory Effect?

(with Klaus Birkelbach)

The Big Fish Little Pond effect (BFLPE) suggests that regardless of their individual ability, students show lower self-confidence in classes with a high average ability – which in turn causes a significant decrease in individuals' school performance. Conversely, the Reflected Glory effect (RGE) hypothesis postulates that upward comparisons lead to a more critical self-evaluation, an enhancement of motivation and thus to an increase in school performance. Our *theoretical* contribution is that we first test whether comparable reference-group effects on teachers' evaluations vary by both student-level achievement and teachers' frames of reference in terms of their grading concepts. Our *methodological* contribution is that we use a cross-classified design where, first, teachers' evaluations are the lowest unit of analysis which is nested in both student- and teacher-level contexts. We then introduce class-level indicators as an additional higher-level unit. Results based on the initial survey of the Cologne High School Panel (1969/70) indicate that i) both class-level socioeconomic status and achievement increase students' probability of obtaining a positive evaluation; ii) this positive effect significantly interacts with student-level achievement; iii) but not with teachers' grading concepts.

1 Introduction

Although there is a non-negligible controversy in educational psychology whether 'grade point average' (GPA) or 'intelligence' has a higher predictive validity for students' later educational course of life, the fact *that* both theoretical concepts are significant is essentially undisputed (Burton and Ramist, 2001; Camara, 1998; Camara and Echternacht, 2000; Camara et al., 2003; Kuncel et al., 2001; Morgan, 1989; Robbins et al., 2004). However, teachers also ground the evaluations of their students on factors other than students' actual achievement[1] – and they may even communicate this to the students

[1]In this paper, the term 'achievement' covers both students' intelligence and their GPA, which we will introduce as two different achievement indicators in the operationalization section.

and their parents. In this study, we aim to shed light on how the emergence of a specific kind of evaluation – i.e. teachers' assessment whom of their 10^{th} class students they believe to be able for academic studies, and whom of them not – might be affected by reference-group effects and their cross-level interactions with both student- and teacher-level variables.

Although the evaluations that have been surveyed in our data, the *C*ologne *H*igh *S*chool *P*anel (*CHiSP*), were never explicitly communicated either to the students or their parents, several studies suggest that even rather implicit teacher assessments may be insinuated in day-to-day school life and thus could affect students' self-concept as well (Brophy and Good, 1974; Brattesani et al., 1984; Good and Brophy, 2003). Therefore, in the long run, even unjustified evaluations by teachers may prove themselves to be true in terms of a *self-fulfilling prophecy* (Merton, 1948; Rosenthal and Jacobson, 1968; Jussim and Harber, 2005; Becker, 2010; Birkelbach, 2011).

While the analysis of student-level predictors of teachers' evaluations has been of interest in educational sociology at least since Alexander et al. (1987), applying the framework of reference-group effects on teachers' evaluations started with Tiedemann and Billmann-Mahecha (2007) and Trautwein and Baeriswyl (2007). In line with these research traditions, the crucial questions in our study are i) to what extent do teachers' evaluations in our data actually depend on reference-group effects *apart from* student-level predictors, and ii) how strongly do potential sizes of suchlike effects vary with both student- and teacher-level variables.

Research analyzing the impact of class- or school-level predictors on students' self-concept and/or achievement usually trades under the name of *Big-Fish-Little-Pond effects* (*BFLPE*; in case of a negative contrast process) or *Reflected Glory effect* (*RGE*; in case of a positive assimilation process) studies. While the initial study of Davis (1966), "The Campus as a Frog Pond"[2], and several sociological follow-ups (Meyer, 1970; Alexander and Eckland, 1975) analyzed contextual-level effects on educational outcomes such as high school attainment or college aspirations, beginning with Marsh and Parker (1984), students' self-concept became the main dependent variable.

In spite of Marsh's (1991) intention to broaden this interimly narrowed focus again by applying the logic of reference-group effects to an extended range of outcome variables[3], there were only few attempts to replicate Marsh's (1991) framework. Thus, more recent studies (Plucker et al., 2004; Rindermann and Heller, 2005) still criticize the focus of most *BFLPE* studies on students' self-concept as being too narrow.

A second disregard in *BFLPE*-alike studies is that although already Davis (1966) stressed the need for an analysis of the underlying educational climates that may contribute to observable differences between schools (or classes; also see Marsh, 1991; Marsh and O'Mara, 2010), only Lüdtke et al. (2005) included an indicator for *teachers' frames*

[2]While Davis (1966) called the observed reference-group effect '*Frog* Pond' effect, we will refer to it as the 'Big-*Fish*-Little-Pond effect' – which is the established term nowadays (Marsh, 1987; Trautwein and Lüdtke, 2005; Dai and Rinn, 2008; Marsh et al., 2008).

[3]These outcomes included students' academic and general self-concept, course-work selection, academic effort, educational and occupational aspirations, school grades, standardized test scores, college attendance, and students' aspirations two years after high school.

of reference (TFR).

Recently, several Swiss and German studies tested how reference-group effects influence the emergence of primary school teachers' transition recommendations (Tiedemann and Billmann-Mahecha, 2007; Trautwein and Baeriswyl, 2007; Gröhlich and Guill, 2009; Milek et al., 2009). However, none of these analyses tested for potential moderating effects of *TFR*s (which is even more surprising than in the case of the classical *BFLPE* studies in the tradition of Marsh and Parker, 1984). Furthermore, in contrast to existing *BFLPE* research with students' self-concept as an outcome (Marsh et al., 1995; Marsh and Rowe, 1996; Marsh et al., 2001; Marsh and Hau, 2003), the above-mentioned studies lack consideration of cross-level interaction terms testing whether class-level achievement generalizes across student ability levels.

Hence, our theoretical contribution is that we extend existing research i) by shedding light on potential reference-group effects concerning the emergence of teachers' 10^{th} class evaluations of their students' expected success at university; and ii) by including measures for *TFR*s and their interaction with our class-level achievement predictors as well as iii) cross-level achievement interaction terms.

Our methodological contribution is that we test this model by means of a more complex multilevel framework where teachers' evaluations are nested in both student and teacher contexts, which are in turn nested in school classes.

In the following sections, we first discuss the main findings from *BFLPE* research, and related evidence about reference-group effects concerning primary school teachers' transition recommendations whereupon we ground our hypotheses regarding 10^{th} class teachers' evaluations.[4] In this context, we especially emphasize the hypothesized counterbalancing effect of *BFLPE* and *RGE* which are supposed to operate simultaneously (Marsh et al., 2000)[5], and we deduce hypotheses about both student-level and teacher-level moderators of class-average achievement effects on teachers' evaluations. In a short paragraph, we provide a brief review of potential student-level predictors of teachers' transition recommendations which we aim to include in terms of covariates. Next, we present our empirical models that result in cross-classified multilevel analyses with three different predictor levels. After the discussion of our findings, we conclude with a brief summary and several recommendations for future analyses.

2 Theory and Hypotheses

In the US context, the analysis of teachers' evaluations has mainly focused on cultural differences due to racial matching or mismatching between students and their teachers (Alexander et al., 1987; Farkas et al., 1990; Downey and Pribesh, 2004). In Germany, there is a great deal of literature on the formation of primary-school teachers' transition

[4] Given that these kind of partly binding evaluations at the end of primary school are most prominent in Switzerland, the Netherlands, and Germany, we will primarily refer to findings from these countries.

[5] Our sample consists of German highest-track students only, and in the original *BFLPE* framework with students' self-concept as an outcome, the latter was found to be *positively* affected by attendance of higher-status schools.

recommendations to their students (see e.g. Arnold et al., 2007; Baumert et al., 2007; Bos and Pietsch, 2004; Ingenkamp, 1977, 1993; Lehmann et al., 1997; Jürges and Schneider, 2006; Pietsch and Stubbe, 2007). These recommendations describe the appropriate secondary school track choice for each student at the end of primary school. Drawing on findings from *BFLPE* research, more recently, reference-group effects on primary-school teachers' transition recommendations have also been analyzed (Trautwein and Baeriswyl, 2007; Tiedemann and Billmann-Mahecha, 2007; Gröhlich and Guill, 2009; Milek et al., 2009). Before we will use results from these studies to deduce reference-group hypotheses regarding 10^{th} class teachers' evaluations in our data, a more extensive discussion of *BFLPE* research in general appears to be reasonable.

2.1 Class-Level Predictors

The well-known study by Davis (1966) is usually referred to as being the first analysis which revealed that students' educational outcomes may be influenced not only by individual-level predictors but by school-level variables as well.[6] Applying the theory of relative deprivation by Samuel A. Stouffer, Davis (1966) found evidence for his hypothesis that there is a negative relationship between school quality – measured by school-average test scores – and both students' grade point average and their college aspirations once students' individual-level ability is controlled for. Shortly after, Meyer (1970) provided a joint analysis of both the positive effect of school-level socioeconomic status (SES) and the negative effect of school-level ability on students' college intentions – whereas Alexander and Eckland (1975) observed that school-level effects on both mid- and long-term educational outcomes are particularly mediated by direct effects on students' educational performance (also see Alwin and Otto, 1977).

Given these more sociological analyses, starting with Marsh and Parker (1984), however, primarily scholars in the field of social psychology became interested in the effect of school-level predictors – which might explain why students' self-concept emerged to be the crucial outcome in reference-group effect research. While negative school-average ability effects on students' self-concept were labeled as *Big-Fish-Little-Pond effects (BFLPE)*, positive effects of school-average social status are denoted as *Reflected-Glory effects (RGE)*.

In the following, we will discuss the main findings from empirical tests of both (competing and/or counterbalancing) hypotheses, and how these may be related to our outcome, teachers' evaluations of students' prospective university success.

2.1.1 Big-Fish-Little-Pond (Or Contrast) Hypothesis

The crucial idea of *BFLPE* is that students may still differ in their academic self-concept even when controlling for their own cognitive abilities due to varying average class-level achievement: A student with given cognitive ability in a low-level learning environment may consider herself to be kind of a *big fish*, i.e. comparably well-positioned, while in

[6] However, Meyer (1970) provides a review of even earlier studies.

a high-level learning environment, she might be equipped with a more negative self-concept. This leads to the paradoxical implication that talented students may develop a lower self-concept when sent to gifted schools, which might in turn have a negative impact on their school achievement due to the reciprocal relationship between self-concept and academic performance (Marsh et al., 2005; Marsh and Craven, 2006).

This basic idea is empirically well-founded (e.g. Marsh and Parker, 1984; Marsh, 1987; McFarland and Buehler, 1995; Zeidner and Schleyer, 1998; Lüdtke et al., 2005; Rindermann and Heller, 2005; Trautwein et al., 2006; Seaton et al., 2010; Thijs et al., 2010)[7], long-lasting (Marsh et al., 2000, 2001, 2007; Marsh and O'Mara, 2010), and generalizable i) across countries (Marsh and Hau, 2003; Seaton et al., 2009), ii) across students' ability levels (Marsh et al., 1995; Marsh and Rowe, 1996; Marsh et al., 2001; Marsh and Hau, 2003), and iii) also to other outcomes apart from self-concept both in education (Marsh, 1991; Marsh and O'Mara, 2010) and beyond (e.g. Chanal et al., 2005; Trautwein et al., 2008).

It is often argued that teachers are influenced by reference-group effects as well (Reuman, 1989; Pallas et al., 1994; Lüdtke et al., 2005). Consequently, regarding the generalizability of *BFLPE* in educational settings, several studies found that students' school grades are indeed affected in this manner (Alwin and Otto, 1977; Marsh, 1991; Trautwein et al., 2006). Several more recent German and Swiss studies applied the idea of reference-group effects on primary school teachers' transition recommendations regarding students' secondary school choice. Trautwein and Baeriswyl (2007) observed a negative effect of school-class ability on teachers' transition recommendations even when controlling for class-level SES (which itself did not reach statistical significance) based on a sample of $N = 741$ 6^{th} grade students in the Swiss canton Fribourg. However, this effect was mediated by teachers' evaluations of both students' academic performance and some motivation-related covariates. Drawing on the Hanover Elementary School Study 2000 ($N = 620$ students), Tiedemann and Billmann-Mahecha (2007) also found that teachers' transition recommendations were negatively affected by class-mean spelling achievement and class-mean cognitive abilities. In their analysis of the German *KESS* (*K*ompetenzen und *E*instellungen von *S*chülerinnen und *S*chülern) data ($N = 11.356$), Gröhlich and Guill (2009) only observed negative effects of students' 4^{th} grade class-mean math ability, but not of class-mean reading ability on teachers' transition recommendations at the end of primary school. However, the former effect was completely mediated by students' individual academic performance. And finally, exactly the same pattern – a significant negative effect of class-level ability that was mediated by students' school grades – was corroborated by Milek et al. (2009) in their analysis of the German PIRLS-E sub-sample. Hence, aside from potential mediators, we assume that a similar reference group effect might affect teachers' evaluations of their students' future academic aptitude, moving teachers to lower their standards in classes with a lower average achievement:

H_1*: In low-achievement classes, students have a higher probability of obtaining a better*

[7] For further research overviews and related discussions see Trautwein and Lüdtke (2005); Dai and Rinn (2008); Dijkstra et al. (2008) and Marsh et al. (2008).

evaluation by their teachers than in high-achievement classes (and vice versa).

2.1.2 Reflected-Glory (Or Assimilation) Hypothesis

Despite strong empirical support for the *BFLPE* (see references above), already Marsh (1987) highlighted that the negative effect of school- or class-level achievement on students' academic self-concept is actually a net effect of both counterbalancing (negative) *BFLPE* and a (positive) *Reflected Glory effect (RGE)* at the same time. While the former usually assumes a mechanism of *deprivation* (Davis, 1966) or a contrast effect (Marsh and Parker, 1984, and ensuing studies), *RGE* expects a mechanism of *assimilation.*

In the eponymous study by Cialdini et al. (1976), the authors demonstrated that students were more likely to identify with their university's football team when the latter was playing successfully. Similarly, it is not unlikely that attending high ability and/or high status classes or schools could also *positively* affect students' self-concept. This is why Marsh (1987) argued that the observed negative *BFLPE* is in fact an effect *net of RGE.*

While the early sociological studies already found positive effects of a school's status on students' educational aspirations (Meyer, 1970; Alexander and Eckland, 1975; Alwin and Otto, 1977), and although Marsh and Parker (1984) and Marsh (1987) controlled for school-level SES[8], later *BFLPE* studies usually disregarded this important contextual-level predictor until it was reintroduced by Marsh et al. (2000). The latter authors were able to juxtapose *BFLPE* and *RGE* by demonstrating that while the effect of school status itself was positive, the effect of school-average achievement on students' academic self-concept turned out to be more negative when school status was controlled for.[9]

Regarding teachers' recommendations as an outcome, only two of the above-cited studies explicitly considered contextual-level SES: First, Tiedemann and Billmann-Mahecha (2007) found that class-level parental educational background – suboptimally measured by teachers' estimates of parental educational aspirations – affect primary school teachers' transition recommendations only at the .10 significance level. However, no explanation about the particular mechanism (which can be assumed to differ from the one implied in conventional *BFLPE* studies with students' self-concept as an outcome) was specified. Second, Gröhlich and Guill (2009) observed that class-level SES only positively affected teachers' transition recommendations in a significant way when class-average

[8]Note that the results of Marsh and Parker (1984) and Marsh (1987) are not really consistent with the above-cited sociological studies and later findings by Marsh et al. (2000), though. Marsh and Parker (1984) observed a *negative* effect of school-level SES on students' self-concept; and also Marsh (1987) reported a (weak) negative effect in their path models. However, the raw bivariate correlation noted in Marsh (1987) was positive. Other notable inconsistencies relate to modeling school classes vs. entire schools as the unit of analysis, and to the usage of *objective* (Marsh and Parker, 1984) vs. *subjective* (Marsh et al., 2000; Trautwein et al., 2009) contextual status measures.

[9]Other attempts of juxtaposing *BFLPE* and *RGE* referred to potentially intervening learning environments (McFarland and Buehler, 1995; Marsh et al., 2001) which will be discussed in the context of potential teacher-level moderators.

ability was controlled for. The fact *that* the effect of class-level SES was significant even though it considerably correlated with class-level student achievement was interpreted as a hint that parents particularly strive for higher track transition recommendations in high-SES classes – which they do in such a way that even low-SES students benefit from it (Gröhlich and Guill, 2009, p. 167).

With regard to the kind of evaluations in our data that were neither made public nor were binding for the students, neither a mechanism of true 'reflected glory' nor the one indicated by Gröhlich and Guill (2009) appears to be plausible. Instead, we suppose that in our case, the mechanism behind the *RGE* hypothesis could be described more precisely as a particular form of a *halo effect* (Thorndike, 1920): In high-SES classes, teachers might project higher parental educational aspirations – independently of whether they are true or just implied by the teachers – on the students from lower strata as well, which induces them to assigning more positive evaluations on average. The opposite relationship should also hold: In low-SES classes, teachers' evaluations might likewise be downwardly biased with regard to higher-SES students as well. Therefore, for our measure of teachers' evaluations we hypothesize:

H_2: In high-SES classes, students have a higher probability of obtaining a better evaluation by their teachers than in low-SES classes (and vice versa).

Although conventional *BFLPE* research explaining students' academic self-concept as an outcome always found negative average achievement effects even when controlling for contextual-level SES – thus regarding the relation between *BFLPE* and *RGE* as *counterbalancing* –, we have good reason to assume that in our case, even a *competing RGE* hypothesis could be warranted:

Similar to the mechanism of a *Halo* effect regarding the hypothesized positive class-level SES effect as outlined above, teachers in high-achievement classes may have a more positive attitude towards the whole class which they project onto the weaker students as well. We particularly assume that since our sample consists of highest-track (i.e. Gymnasium) students only, teachers could be upwardly biased *a priori* – which is why the *gross* effect of academic achievement on teachers' evaluations could also be positive:

H_3: In high-achievement classes, students have a higher probability of obtaining a better evaluation by their teachers than in low-achievement classes (vice versa).

2.2 Interaction Effects

While the preceding hypotheses regarded teachers' evaluations as an outcome, we additionally test for *cross-level interaction effects* (Hox, 2010, pp. 31-36, 63-69; also called *slopes-as-outcome* approach; see Bryk and Raudenbush, 1992, p. 21), i.e. whether the effect of class-average achievement on teachers' evaluations varies by levels of variables on a different unit of analysis. To be precise, we will look at this potential variance by both student and teacher level predictors.

Student-level moderator As a first theoretical advancement to the existing literature on teachers' evaluations, we postulate that the effect of class-mean achievement on teachers' evaluations varies by students' *individual* ability. Regarding the classical *BFLPE*, most studies found that the effect of class mean achievement was generalizable across different student ability levels. For instance, Marsh et al. (1995) were not able to detect a significant interaction effect, while Marsh and Rowe (1996) did find a significant, but rather small interaction effect. In contrast, Marsh et al. (2001) and Marsh and Hau (2003) found no significant interaction term.

However, none of the above-cited reference-group effect studies of teachers' recommendations tested for cross-level interactions. Just as Coleman and Fults (1985) report that students in the upper level of a class' ability distribution are less affected by *BFLPE*, we assume that the higher a student's achievement, the less pronounced reference-group effects on teachers' evaluations will be: High achievers might be more visible in the classroom than low achievers – which is why students' individual achievement will be a more important criterion for teachers in the case of high achievers than in the case of low achievers.[10]

An alternative interpretation (or underlying mechanism) of this negative interaction effect would be that the impact of student-level achievement on teachers' evaluations is weaker in high-achievement classes, since each student would have to be equipped with a comparatively higher aptitude to stand out from her classmates, or a teacher might (unconsciously) refer to alternative criteria (e.g. parental backgrounds) to a greater extent when evaluating her students. We choose to adhere to the convenient framework of a moderation of class-level predictor effects by student-level covariates. Therefore, our hypothesis nonetheless reads:

H_4: The effect of class-level achievement on teachers' evaluations decreases with students' individual achievement.

Teacher-level moderator Apart from the above-mentioned studies on student achievement, there is little evidence regarding moderators of reference-group effects.[11] Thus, in the study at hand, we first test for the variance of reference-group effects on teachers' evaluations by teachers' grading concepts.

Drawing on Rheinberg (1980) and his distinction between more social and more individualist reference standards of teachers, Lüdtke et al. (2005) interacted indicators of those two concepts with the effect of class-average math achievement on students' math self-concept.[12] Although none of these interaction terms were statistically significant,

[10] The *Halo* effect was observed to vary with the ratee's relative visibility (see Kozlowski and Kirsch, 1987).

[11] Dai and Rinn (2008) make the point that previous research has lacked a "specification of contexts where the *BFLPE* is more or less likely to occur" (p. 297). A recent exception is the study by Thijs et al. (2010) that found students' perceived relative class position and both a school's class size and gender composition to moderate the *BFLPE*.

[12] Concerning self-concept as an outcome, McFarland and Buehler (1995) observed that the *BFLPE* is stronger for students with a more individualist cultural heritage than for those with a more

there are good reasons to assume that the underlying mechanism is actually more likely to affect the impact of reference-group effects on teachers' evaluations rather than on students' self-concept: Teachers with a more individualist reference standard might be less prone to upward or downward movements of their evaluations depending on class-level achievement than teachers who reported a more relational reference standard. This is why we hypothesize:

H_5: The relationship between class-level achievement and teachers' evaluations (BFLPE/ RGE) varies with teachers' reported grade concepts (relational vs. individualist).

2.3 Student-Level Covariates

We include indicators for students' achievement, their social backgrounds, and their aspirations as student-level covariates. Regarding individual *achievement*, in the US, standardized tests such as *ACT* or *SAT* are a broadly used – albeit critically debated – tool in order to select college students after high school (Burton and Ramist, 2001; Kuncel et al., 2001; Morgan, 1989; Robbins et al., 2004). In spite of the fact that German pedagogic researchers also acknowledge students' intelligence to be the most important single predictor for their educational success (e.g. Ingenkamp, 1993, p. 73), in the German educational system, there are still no serious attempts to ground students' educational tracking on test results. Although more recent studies such as *Progress in International Reading Literacy Study (PIRLS)* 2001 and 2006 (Bos et al., 2004; Arnold et al., 2007) as well as the German studies *LAU: Aspekte der Lernausgangslage und der Lernentwicklung* (Lehmann et al., 1997) and *KESS 4: Kompetenzen und Einstellungen von Schülerinnen und Schülern – Jahrgangsstufe 4* (Bos and Pietsch, 2004) showed that teachers' transition recommendations and students' test results only partly overlap, a virtually linear positive relationship between students' language comprehension, their reading abilities, their ability to extract information, and their math scores as independent variables, and teachers' recommendations as the outcome nonetheless holds (Lehmann et al., 1997). This is what we would also expect concerning teachers' evaluations.

Being a more visible achievement criterion to teachers (see e.g. Alexander et al., 1987, p. 667), students' grades have a higher predictive power in explaining teachers' transition recommendations than students' intelligence (Ingenkamp, 1993; Kristen, 2006). In *PIRLS* 2006, about 69% of the total variance in teachers' recommendations can be explained by both students' language and math grades (while the effect of grades in language was a bit more pronounced than in math; see Arnold et al., 2007, p. 283). Since school grades, though distorted by well-known deficiencies, are empirically well-tried predictors of teachers' recommendations, we consider them to be a relevant predictor of teachers' evaluations as well.

collectivist one. Similarly, Marsh et al. (2001) found that *BFLPE* was more negative for West German students than for East German students shortly after the reunification. This result could be replicated by Seaton et al. (2009) based on the PISA data with 41 countries showing that *BFLPE* significantly interacts with countries' degree of individualism.

With regard to students' *social backgrounds*, once more research based on *PIRLS* demonstrated that even when controlling for students' (general) cognitive and (more specific) reading abilities, students from service-class parents have a 2.6 times higher probability of obtaining a recommendation for *Gymnasium* compared to working-class offsprings (Bos et al., 2004, p. 213; Arnold et al., 2007, p. 289). Apart from primary social class gross effects on teachers' evaluations due to actual achievement differences, following Bourdieu and Passeron (1990), service-class students would still obtain better evaluations by their teachers since they dispose of the symbolic codes which are implicitly demanded in classroom communication.

An additional secondary effect of social inequality may occur via the variation of students' *aspirations* by their parents' social class due to social downward mobility avoidance (also known as the mechanism of *relative risk aversion*; see Boudon, 1974; Goldthorpe, 1996; Erikson et al., 2005; Müller-Benedict, 2007). We assume that secondary effects of social inequality in terms of differing educational aspirations not only influence actual transition decisions (Meulemann, 1979; Breen and Goldthorpe, 1997; Esser, 1999; Erikson et al., 2005; Maaz et al., 2006; Stocké, 2007) – but also the preceding teachers' recommendations since they can be thought to be part of the communication between students (or their parents) and teachers. Apart from verbal communication, higher aspirations might also influence certain forms of students' behavior as what is typically referred to as the *habitus* of higher-class students can be supposed to match to particular norms and expectations of the teachers (Bourdieu, 1986).

Finally, analyses control for students' gender. Figure 1 summarizes all theoretical concepts on different levels of analysis.

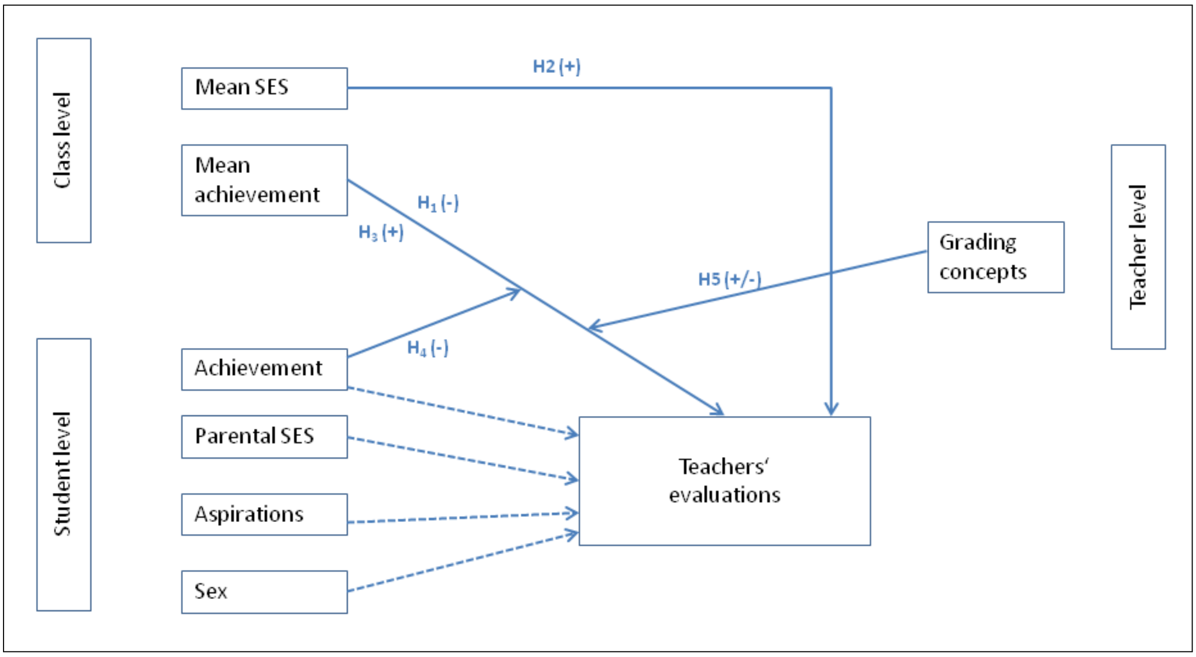


Figure 1: Summary of hypotheses.

3 Data & Methods

3.1 Data

For our analysis, we use the initial survey of the *Cologne High School Panel (CHiSP)* from 1969/1970. Funded by the Federal State of North-Rhine Westphalia, the Research Institute of Sociology at the University of Cologne surveyed $N = 3.38510^{th}$ class students (nested in $N = 120$ school classes) at the highest German school track – *Gymnasium* – in North-Rhine Westphalia. Part of the student questionnaire (Gesis-No. ZA600) covered issues such as students' interest in school (and in certain subjects), their achievement, their future educational and occupational plans, their social backgrounds and their attitudes towards school. Remarkably, students also took part in an intelligence structure test (*IST;* Amthauer, 1957). In addition to the student questionnaires, students' parents (Gesis-No. ZA639; $N = 2646$) and their teachers (Gesis-No. ZA640; $N = 1701$) were surveyed as well.[13]

3.2 Variables

3.2.1 Dependent Variable

In two questions, teachers were asked to assess whom of their students they consider to be able to make the transition to university, and whom of them not. In a recent study based on the same data, Becker and Birkelbach (2010) analyzed the impact of reference-group effects on the evaluations of class-teachers only. Here, 751 students were considered by their class teachers as being 'able', and 616 students as being 'not able', while 1060 students did not receive any evaluation at all. Becker and Birkelbach (2010) provided the argument that overall, teachers feel certain enough in their evaluations for high and low achieving students to give an explicit prognosis, while they would rather abstain from that for 'average' students. This claim was supported by an analysis of students' mean intelligence scores in all three categories: For students without an explicit evaluation, the arithmetic mean of their intelligence is about $\overline{IST}_{TE_0} = 109.6$, for students with a positive evaluation $\overline{IST}_{TE_+} = 114.7$, and for students with a negative evaluation $\overline{IST}_{TE_-} = 107.8$.[14] We aim to apply a similar logic here, but in contrast to Becker and Birkelbach (2010), we do not restrict our analyses to class teachers' evaluations. Instead, we take into account that each student potentially received an evaluation by several teachers, and each teacher potentially evaluated several students. In total, there are $N = 17.626$ evaluations in our data, of which 4.363 evaluations are positive, and 3.607 evaluations are negative. Due to space constraints, and to facilitate computations, our dependent variable is dichotomized into a positive vs. a negative evaluation.

[13] As a further amendment of the initial survey, an additional survey was administered to the school principals (Gesis-No. ZA996; $N = 68$). Later, the former students were re-surveyed in 1984/85 (Gesis-No. ZA1441, $N = 1987$), in 1996/97 (Gesis-No. ZA4228; $N = 1596$) and 2010 ($N = 1301$, no Gesis-No. available yet).

[14] This trend is similar for students' school grades, and also holds in a multinomial logistic regression analysis with additional covariates (see Becker and Birkelbach, 2011, appendix, Tables B and C).

3.2.2 Independent Variables

Student-level Students' *intelligence* was measured along four sub-scales of the intelligence structure test (`ist`) that the students took part in. We used the group-mean-centered[15] total score of all four dimensions. Students' *school achievement* was operationalized by their group-mean-centered grade point average (`gpa`). For their *social backgrounds*, we draw on a three-dimensional index of parental social class (`ses`) based on parents' occupational status and situs that has already been added to the data (1 'lower working class'; 2 'upper working class'; 3 'lower middle class'; 4 'middle class'; 5 'upper middle class'; 6 'upper class'; see Meulemann, 1979, p. 49). And finally, students' *aspirations* (`aspir`) were measured by students' answer to the question if they considered 'Abitur' to be necessary to reach their aim in life (0 'not necessary / no aim in life'; 1 'useful or necessary'). Apart from these indicators, analyses also control for students' gender (`sex`; 0 'female'; 1 'male').

Class-level To test for *BFLPE* (contrast effects) and *RGE* (assimilation / halo effects), respectively, we first computed two measures of class-level achievement and controlled for class-level SES. Additional to the grand-mean-centered class-level means of students' intelligence (`IST`), we computed grand-mean-centered class-average grades (`GPA`).[16]

Second, we also computed grand-mean-centered class-level `SES` to measure potential *RGE*s that might counterbalance *BFLPE*s. For school classes with missing values on the student level, we computed the grand-mean-centered class-level means of all non-missing observations.

Teacher-level As mentioned before, teachers were asked which criteria they ordinarily use for their grading decisions. Apart from a list of predefined categories such as written test scores or pedagogic incentives, teachers could also name additional criteria as a first or a second choice. 21 (≙3.92%) out of 536 teachers named an individualist grading concept as a second choice, and 47 (≙8.77%) considered it to be their first choice. Contrarily, 7 teachers (≙1.31%) mentioned a relational grading concept as a second choice, and 26 teachers (≙4.85%) considered it to be their first choice. We created two ordinal variables (`TFR(ind)` and `TFR(rel)`) which were assigned a 1 if teachers explicitly

[15] Hox strongly recommends to use group-mean centering techniques only if they are supposed to represent a specific hypothesis since the entire model is changed and cannot be re-transformed to the raw scores by means of simple algebraic transformations unlike with grand-mean centering (Hox, 2010, p. 68f.). Note that the analysis of frog-pond effects entails suchlike hypotheses where group-mean centering is an appropriate specification (Kreft et al., 1995; Hox, 2010, p. 68f.).

[16] Although not an objective measure of student ability as Marsh et al. (2008) consider it to be one of the three minimal conditions for 'true' *BFLPE* studies, we have good reasons also to include class-level grade point average as a contextual-level predictor. A simple empirical reason is that Tiedemann and Billmann-Mahecha (2007) observed significant effects of class mean spelling performance on teachers' transition recommendations, and we want to ensure that our study is as comparable as possible with preceding analyses. A more substantive theoretical reason is that teachers can be assumed to exchange views about student aptitudes in the faculty room (Mechtenberg, 2009), and thus a teacher evaluating a certain student might be influenced by opinions of other teachers.

named one of the respective motives as a second choice, a 2 if teachers named it as a first choice, and 0 otherwise. Table A lists the summary statistics of all independent variables and table B their inter-correlations.[17]

3.3 Method and Models

Since our hypotheses cover different levels of analysis, we have to bear in mind the nestedness of our data by *multilevel* or *hierarchical modeling* (Bryk and Raudenbush, 1992; Goldstein, 1995; Hox, 2010; Snijders and Bosker, 1999). Figure 2 illustrates our approach to account for this hierarchy: We consider teachers' evaluations to be the lowest unit of analysis, and to be simultaneously nested in both teacher and student contexts. Thus, for teachers and students we estimate a *cross-classified* multilevel model. The reason why we do so is that unlike an existing study based on the same data (Becker and Birkelbach, 2010), we do not use evaluations of class teachers only – but of all teachers that completed the underlying questionnaire. Hence, each student potentially obtained evaluations by multiple teachers; but reversely, each teacher potentially evaluated multiple students, and so we nested teachers' evaluations in both student and teacher contexts. Furthermore, we assume teachers and students to be nested in school classes.

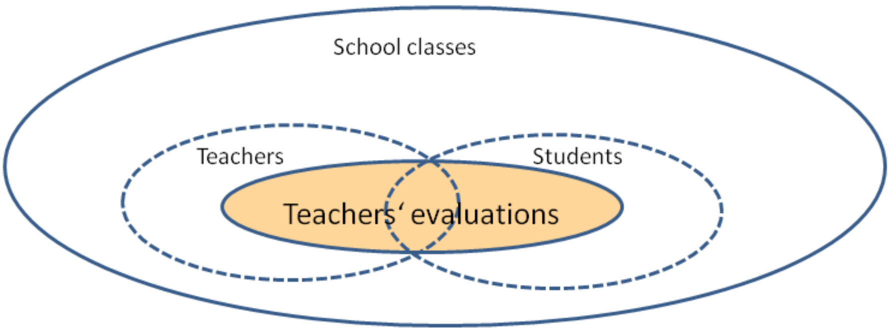


Figure 2: Data structure.

Because our dependent variable is dichotomous, we estimate a logistic multilevel model. We use the `lme4` package in *R* (Bates and Maechler, 2009)[18] to estimate the following models: First, we discuss the null model without any predictors to give an intuition about the variance parameters at different levels of analysis. Second, we present the minimum *BFLPE* model including each indicator of class-level achievement and both its corresponding and its complementary individual-level term (*models 1a-1f*). As

[17]Although the distribution of our *TFR* indicators is quite skewed, we favor an analysis of imperfectly-distributed variables over abstaining from the corresponding models completely.

[18]The `lme4` package has the advantage that it uses Laplace approximation which has been shown to be better than penalized quasi-likelihood methods (which result in downwardly-biased estimates) and at least as precise as Gauss-Hermite quadrature methods but computationally faster (see Raudenbush et al., 2000, for a simulation study). Another pleasant feature of `lme4` is that it will automatically choose the correct nesting structure given each cluster unit has its unique identification number as is the case in our data.

the potential variance of contextual-level effects by student-level covariates has always been an issue in frog-pond research – though, to the best of our knowledge, was never analyzed in the context of reference-group effect research on teachers' evaluations –, we then test whether potential contextual-level achievement effects vary by students' individual achievement level. Therefore, in *models 2a-2d*, we add cross-level interaction terms between class-level and student-level achievement, and we test whether observed class-level achievement effects and cross-level interaction effects remain stable with additional student-level controls (*models 3a-3f*).

To test for moderation effects of teachers' frames of reference, we start again with the minimum *BFLPE* model (first without cross-level achievement interaction terms) and add two dummy indicators for teachers' more individualistic or more relational grading concepts, respectively, as well as their cross-level interaction terms with both class-level intelligence and GPA (*models 4a-4h*). We then test whether the observed effects remain stable with controls for cross-level achievement interaction effects (*models 5a-5h*) and individual-level student covariates (*models 6a-6l*).

In a final step, we analyze whether estimated reference-group effect sizes change when class-average SES is added to the equation. In *models 7a-7g*, we subsequently introduce class-mean SES, both class-mean achievement and student-level achievement indicators and their cross-level interaction terms. And in *models 8a-8h* we control for student-level covariates.

4 Results

Null model The null model is a "one way ANOVA with random-effects" (Bryk and Raudenbush, 1992, p. 17) and as such, in multilevel models it disentangles the variance components on different level of analyses. Moreover, it is crucial to determine the *Intra-class Correlation Coefficient (ICC)* in order to assess the *amount* of variance on a given cluster level compared to the total outcome variance. In our case, students' probability of obtaining either a positive or a negative evaluation may vary according to student-level, teacher-level, and class-level factors, while a residual term would point to the amount of (lowest-level) variance of teachers' evaluations that cannot be explained on the specified cluster levels. Since we are fitting logistic multilevel models, this lowest-level variance has to be fixed to the known variance of the logistic distribution of $\pi^2/3 \approx 3.29$ (Hox, 2010, p. 135).

Looking at the variance components in the empty model, we can compute the following ICCs:

$$ICC_{student} = \frac{13.1029}{13.1029 + 1.6602 + 2.5673 + \pi^2/3} \approx .635 \tag{1}$$

$$ICC_{teacher} = \frac{1.6602}{13.1029 + 1.6602 + 2.5673 + \pi^2/3} \approx .081 \tag{2}$$

$$ICC_{class} = \frac{2.5673}{13.1029 + 1.6602 + 2.5673 + \pi^2/3} \approx .125 \tag{3}$$

As can be seen, most of the variance of teachers' evaluations can be expected to be explained by student-level predictors, while class-level predictors come in second. The crucial question for the following analyses is whether we will be able to make a substantive contribution by including teacher-level predictors, since on this level, the lowest part of outcome variance can be expected. However, since both a likelihood ratio test (Vuong, 1989) and AIC as well as BIC comparisons[19] suggest that the null model including a random intercept for the teacher level fits the data significantly better than a corresponding null model ($\chi^2 = 246.98, Pr(> Chisq) < 2.2e^{-16}, df = 1$), we still see good reasons to allow for a suchlike effect.

Another measure for the prospective explanatory power of a multilevel model is the design effect d proposed by Muthen and Satorra (1995), which considers the average cluster-unit group size.[20] In our case, we have three separate measures for the design effect, i.e. one for each ICC:

$$d_{student} = 1 + (3.572 - 1) \cdot .635 \approx 2.633 \tag{4}$$

$$d_{teacher} = 1 + (19.109 - 1) \cdot .081 \approx 2.467 \tag{5}$$

$$d_{class} = 1 + (90.289 - 1) \cdot .125 \approx 12.161 \tag{6}$$

If $d > 2$ at a given cluster level, a multilevel model is supposed to be justified (Muthen and Satorra, 1995; Maas and Hox, 2005) – which is the case on each level of analysis in our data.

Class-level model: achievement In line with the methodological paradigm of *BFLPE* research, we first fit models including only class-level intelligence and GPA. As can be seen from *model 1a* (table 1), the higher the grand-mean centered class-level intelligence, the higher students' probability of obtaining a positive evaluation by their teachers; and, similarly, the higher the grand-mean centered class-level GPA, the higher students' probability of obtaining a positive evaluation by their teachers (*model 1b*).

This is in line with our interpretation of the *RGE* hypothesis in terms of a Halo effect (H_3) in that teachers tend to project their comparably higher expectations in high-achievement classes onto lower-achieving students as well (or simply higher their reference standards). Contrarily, the *BFLPE*- or contrast hypothesis H_1 has to be rejected. This finding contrasts the majority of *BFLPE* studies with self-concept as an outcome (cf. reviews by Trautwein and Lüdtke, 2005; Dai and Rinn, 2008; Dijkstra et al., 2008; Marsh et al., 2008) as well as reference-group effect research with primary

[19] *A*kaike's *I*nformation *C*riterion (*AIC;* Akaike, 1974) judges the fit of a given model by relating its maximized value of the likelihood function L to the number of parameters k in the model: $AIC = 2k - 2ln(L)$. The *B*ayesian *I*nformation *C*riterion (BIC) Schwarz (1978) is related to the AIC but introduces a penalty for the number of parameters in the model. The formula for the BIC reads $-2 \cdot ln \cdot p(x \mid k) \approx BIC = -2 \cdot ln \cdot L + k \cdot ln(n)$. For further information see DeLeeuw (1992); Raftery (1995); Burnham and Anderson (2002) as well as Burnham and Anderson (2004).

[20] The formula for the design effect reads $d = 1 + (average\ group\ size - 1) \cdot ICC$.

school teachers' transition recommendations as an outcome (Tiedemann and Billmann-Mahecha, 2007; Trautwein and Baeriswyl, 2007; Gröhlich and Guill, 2009).

Though student-level intelligence significantly predicts teachers' evaluations, it neither alters the estimate nor the z-value of class-level intelligence. For class level GPA, both estimate and z-value tend to decrease a bit when student-level GPA is controlled, but both size and significance level of the class-level coefficient remain relatively stable. Thus, the Halo effect we found persists in that students in high-achievement classes still obtain better evaluations by their teachers even when their individual achievement is controlled for.

The models controlling for the complementary achievement indicator also allow the conclusion that student-level GPA is a more powerful predictor of teachers' evaluations than student-level intelligence, as introducing the former leads to a downsize of class-level intelligence, while introducing the latter does not have the same effect on class-level GPA.

All tables report the conditional ICCs of the fitted models as well as Nagelkerke's R^2, which is defined as follows for logistic multilevel models (cf. Hox, 2010, p.135):[21]

$$\frac{R^2 = R^2_{CS}}{1 - exp(-Deviance_{Null}/n)} \tag{7}$$

with R^2_{CS} as Cox and Snell R-square defined as

$$\frac{R^2_{CS} = 1 - exp(Deviance_{model} - Deviance_{null})}{n} \tag{8}$$

Regarding the R-square in table 1, we can see that including only class-level achievement does not substantively improve the model fit. By adding the corresponding student-level term, an R^2 of .13 can be achieved for the class-level intelligence model, and an R^2 of .35 for the class-level GPA model. And with controls for both student-level intelligence and GPA, the model fit improves towards an R^2 of around .37 for both sub-models.

Thus, although both contextual-level achievement predictors are positively significant even when controlling for their respective student-level counterparts, teachers still more strongly rely on student-level achievement as a criterion for their evaluations than they fall prey to a Halo effect.

Also note that a visible reduction in student-level variance is only achieved when student-level GPA is added to the model, which once again indicates the prevalence of this particular achievement indicator.

In table 2, we add cross-level interaction effects between both indicators of class-level achievement and their respective student-level indicators. For each of the models, we estimate both a random slope and a fixed slope model for the student-level variable whereupon a cross-level interaction is performed, and since these two models are nested,

[21] Although Tabachnick and Fidell (2007) warn that these pseudo R-square statistics should not be interpreted as 'explained variance' in its true sense (p. 446f.), according to Hox (2010), they can still be used to assess the basic fit of the underlying model (p. 135). However, note that the pseudo R-squares tend to underestimate the actual model fit compared to the OLS R-square (Long, 1997, pp. 102ff.).

Table 1: Multilevel analysis of teachers' evaluations: Minimum *BFLPE* model

		Model 1a		*Model 1b*		*Model 1c*		*Model 1d*		*Model 1e*		*Model 1f*	
		Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value
	intercept	0.63***	3.82	0.59***	3.39	0.71***	3.8	0.56**	3.11	0.59***	3.46	0.56**	3.12
student	ist					0.17***	18.01			0.06***	8.49	0.06***	8.46
level	gpa							7.35***	30.92	6.96***	29.74	6.95***	29.73
class	IST	0.2***	5.42			0.21***	5.03			0.17***	4.61		
level	GPA			3.64**	3.2			3.28**	2.82			3.23**	2.77
	student-level	0.663		0.647		0.654		0.455		0.449		0.438	
ICC	class-level	0.091		0.108		0.105		0.15		0.139		0.158	
	teacher-level	0.084		0.083		0.094		0.173		0.179		0.175	
	total	7968		7968		7627		7626		7602		7602	
	student	2903		2903		2774		2774		2764		2764	
N	class	119		119		119		119		119		119	
	teacher	520		520		520		520		520		520	
	AIC	7817		7833		7173		5890		5769		5781	
fit	BIC	7852		7868		7215		5932		5817		5830	
measures	Nagelkerke's R^2	0.005		0.002		0.131		0.352		0.372		0.37	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

we can evaluate by means of a likelihood-ratio test (Vuong, 1989) whether the random-slope model fits the data significantly better than a fixed-slope model. In none of the cases does the underlying χ^2 value get close to a conventional level of statistical significance, which is why the estimates of table 2 are consistently based on fixed slope models.

In *model 2a* and *model 2b*, the cross-level interaction effect is not significant regardless of whether student-level GPA is controlled for (as in *model 2b*) or not. Contrarily, *model 2c* reveals a significant cross-level interaction between class-level and student-level GPA – indicating that the effect of student-level GPA on teachers' evaluations increases in classes with higher mean GPA (or that the positive effect of class level GPA on teachers' evaluations is stronger for students with a higher GPA). Controlling for student-level intelligence does affect the significance value of the cross-level interaction term (*model 2d*); however, note that the underlying z-value of *model 2d* is still very close to the critical value of 1.96.[22]

Since even an insignificant interaction term – as in the case of class-mean average IST – does not indicate that both estimate and significance level of a given coefficient do not vary by the categories of another one (see Brambor et al., 2006, p. 70), we computed the marginal effects of our interaction terms as well as their corresponding conditional z-values.[23]

However, also according to these statistics, no notable variance of the class-mean ability effect on teachers' evaluations by student ability levels can be detected. Figure 3 illustrates the difference between the two cross-level ability interaction terms for the minimum, the 25^{th}, 50^{th} and 75^{th} percentile, as well as the maximum for the respective student-level variable.[24] While for the `IST*ist` interaction, marginal differences in the slope variance without controls for student-level GPA are canceled out when this control goes into the model (upper line of the figure), there is still a considerable variance regarding the `GPA*gpa` interaction even with controls for student-level intelligence (lower line of the figure). To be precise, the marginal effect of class-average intelligence on the probability of obtaining a positive teacher evaluation varies between .11 and .31 for the plotted percentiles of student intelligence before controlling student-level GPA, and

[22] The significant cross-level interaction between student-level and class-level GPA remained stable even when controlling for class-level intelligence and its interaction with student-level intelligence (not shown, available upon request). However, in order to avoid convergence problems in the more complex models following in the next paragraphs, we always included only one class-level predictor at a time.

[23] The marginal effect of an interaction term is computed by the first partial derivative of one of the variables involved, which for a simple two-level case yields: $\frac{\Delta \widehat{Y_{ij}}}{\Delta \widehat{X_{1ij}}} = \gamma_{10} + \gamma_{11} Z_{1j}$ with the corresponding standard error $s.e.(\gamma_{10} + \gamma_{11} z_{1j}) = \sqrt{var(\gamma_{10}) + z_{1j}^2 \cdot var(\gamma_{11}) + 2 \cdot z_{1j} \cdot cov(\gamma_{10}, \gamma_{11})}$ (with z_{1j} being a specific value of Z_{1j}; see Friedrich, 1982; Brambor et al., 2006).

[24] We used the `lmer.PlotInt.fnc` in the `languageR` package to plot the interaction effects fitted by `lme4`. Note that the percentiles for which the marginal effects are plotted refer to the distribution of the student-level covariate from the fitted models (net of missing values) and not to its distribution in the raw data. This is why the actual values of the percentiles differ within the plots and compared to the summary statistics in table A (appendix).

between .13 and .21 when the latter variable goes into the model (without any differences in significance regarding the conditional z-values in both cases). Contrarily, the marginal effect of class-average GPA on the probability of obtaining a positive teacher evaluation varies between -.84 and 8.66 for the plotted percentiles of student-level GPA before controlling student-level intelligence, and between -.72 and 8.39 when the latter variable goes into the model. In both cases, the marginal effect of class-average GPA is not significant for the minimum of student-level GPA of -1.35 ($z \approx 1.55$) but significant for the other percentiles (with proportionally increasing conditional z-values).

Hence, while in our hypothesis H_4 we postulated that the effect of class-average achievement should decrease with higher levels of individual student achievement, it is just the other way round: The better a student's grade point average, the more she benefits from a Halo effect in terms of teachers' 'upgrading' their evaluations in classes with a high mean value of GPA. This is in contrast to the study by Marsh and Rowe (1996) who found a small negative interaction effect (though with students' self-concept as an outcome). However, as mentioned, we cannot find a suchlike effect for both student and class-level intelligence – which is in line with Marsh et al. (1995, 2001) and Marsh and Hau (2003) who also failed to observe a significant interaction effect.

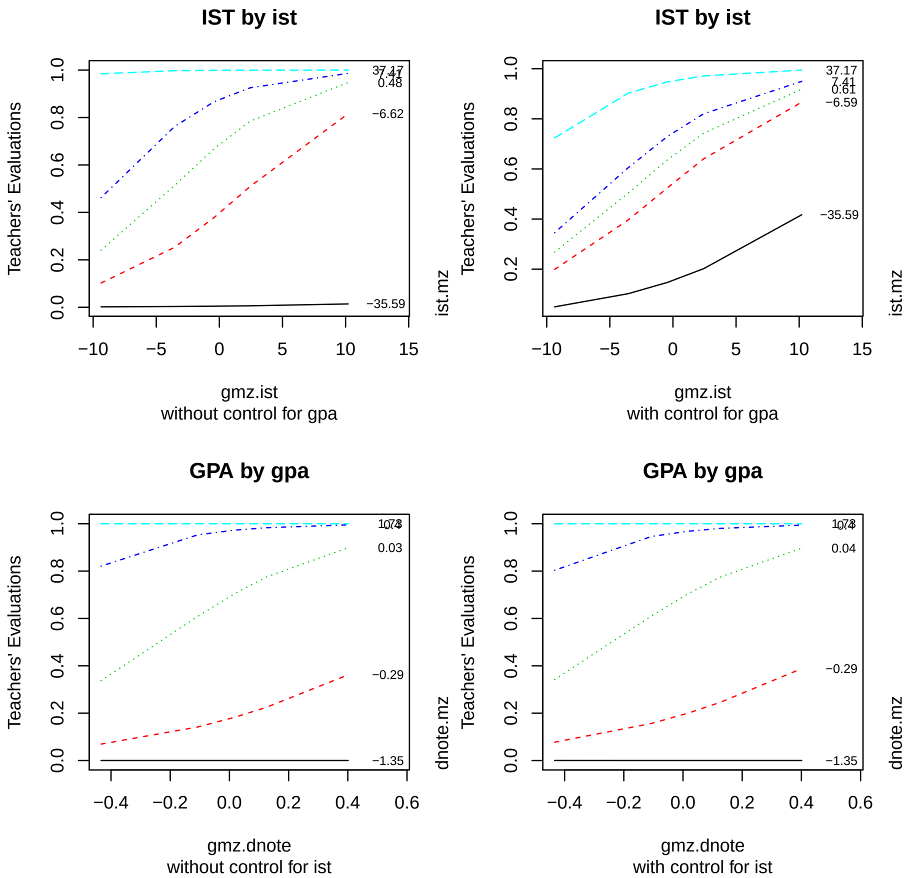


Figure 3: Marginal effects of cross-level ability interaction terms.

Although the cross-level interaction effect between class-level and student-level GPA on teachers' evaluations is statistically significant, regarding the R-square terms, these models do not fit the data remarkably better than *models 1d* and *1f* (table 1) without such an effect.

In a third step, we tested whether the class-level effects obtained in table 1 as well as the cross-level interaction effect observed in table 2 remained stable with additional student-level controls. Since in this series of models cross-level interactions are involved as well, regarding the appropriateness of random-slope models we proceed similarly to our approach in table 2. Again, the estimates in table 3 are fixed-slope models since their random-slope counterparts did not fit the data significantly better.

In *model 3a*, we add parental social class in the equation, which is positively associated with teachers' evaluations, but does neither alter the significant effect of class-level intelligence nor its insignificant cross-level interaction term with student-level intelligence (more precisely, the z-statistic of the interaction term *decreases* with controls for social class). Similar trends hold for an additional term of students' aspirations (*model 3b*), whereas we find a positive effect for male students (*model 3c*).

By content, these results indicate that the higher the social class of students' parents, and the higher students' aspirations, the higher the probability of obtaining a positive evaluation – which is in line with both analyses of primary school teachers' transition recommendations as an outcome (Bos et al., 2004; Arnold et al., 2007; Tiedemann and Billmann-Mahecha, 2007; Trautwein and Baeriswyl, 2007; Gröhlich and Guill, 2009) and a structural equation model based on the same data (Becker and Birkelbach, 2011). Similarly, male students have a higher probability of obtaining a positive evaluation than female students – which contradicts a recent study based on primary school teachers' transition recommendations that found the opposite gender effect (Gröhlich and Guill, 2009).

Models 3d-3e repeated the stepwise procedures from *models 3a-3c* but included class-level GPA and its interaction with student-level GPA instead of class-level intelligence. Regarding the student-level covariates and the 'main effect' of class-level GPA, no substantive difference compared to *models 3a-3c* can be observed. Contrarily, the interaction term between student-level and class-level GPA is only significant with controls for parental social class but not with additional controls for students' aspiration and their gender.

Apparently, a substantive part of the variance of the class-level GPA effect on teachers' evaluations that could be explained by differences in student ability is partly due to differences in students' aspirations. However, since the z-statistics are again not that far from the threshold of 1.96, there is still evidence that the relationship between class-level GPA on teachers' evaluations is not completely insensitive to students' individual achievement in terms of GPA.

Another result from table 3 is that in both *models 3c* and *3f*, the effect of class-level achievement tends to diminish slightly when students' gender is controlled for. We will come back to this finding in our discussion section. Though all of them significant, only minor gains in model fit can be yielded by including the additional student-level covariates.

Table 2: Multilevel analysis of teachers' evaluations: *BFLPE* model + cross-level interactions

		Model 2a		*Model 2b*		*Model 2c*		*Model 2d*	
		Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value
student level	intercept	0.71***	3.8	0.6***	3.47	0.56**	3.14	0.57**	3.15
	ist	0.17***	17.93	0.06***	8.49			0.06***	8.47
	gpa			6.96***	29.74	7.3***	30.82	6.91***	29.63
class level	IST	0.21***	5.01	0.17***	4.62				
	GPA					3.32**	2.85	3.27**	2.81
cross-level interactions	IST*ist	0	1.3	0	0.67				
	GPA*gpa					3.09*	1.97	2.96	1.93
ICC	student-level	0.654		0.449		0.452		0.435	
	class-level	0.105		0.138		0.151		0.159	
	teacher-level	0.094		0.178		0.173		0.176	
N	total	7627		7602		7626		7602	
	student	2774		2764		2774		2764	
	class	119		119		119		119	
	teacher	520		520		520		520	
fit measures	AIC	7173		5770		5887		5779	
	BIC	7222		5826		5936		5834	
	Nagelkerke's R^2	0.131		0.372		0.354		0.37	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table 3: Multilevel analysis of teachers' evaluations: *BFLPE* model + student-level covariates

		Model 3a		Model 3b		Model 3c		Model 3d		Model 3e		Model 3f	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
student level	intercept	0.59***	3.45	0.15	0.73	-0.18	-0.77	0.56**	3.13	0.11	0.52	-0.39	-1.7
	ist	0.07***	8.66	0.06***	8.46	0.06***	8.36	0.06***	8.64	0.06***	8.43	0.06***	8.28
	gpa	6.86***	29.56	6.77***	29.1	6.83***	29.03	6.81***	29.44	6.72***	28.98	6.81***	28.94
	ses	0.28***	4.46	0.25***	3.98	0.25***	4.02	0.28***	4.5	0.25***	4.01	0.26***	4.06
	aspir			0.65***	3.9	0.62***	3.74			0.66***	4.03	0.62***	3.77
	sex					0.64*	2.57					0.96***	3.99
class level	IST	0.17***	4.63	0.17***	4.45	0.14***	3.79						
	GPA							3.26**	2.8	3.23**	2.83	3.83***	3.58
cross-level interactions	IST*ist	0	0.73	0	0.7								
	GPA*gpa					0	0.77	3.11*	2.04	2.92	1.92	2.81	1.84
ICC	student-level	0.443		0.446		0.45		0.428		0.433		0.451	
	class-level	0.142		0.136		0.128		0.162		0.153		0.123	
	teacher-level	0.179		0.179		0.18		0.176		0.177		0.183	
N	total	7588,		7552,		7552,		7588,		7552,		7552,	
	student	2759;		2747;		2747;		2759;		2747;		2747;	
	class	119		119		119		119		119		119	
	teacher	520;		519;		519;		520;		519;		519;	
fit measures	AIC	5742		5709		5704		5750		5716		5703	
	BIC	5804		5778		5780		5812		5785		5779	
	Nagelkerke's R^2	0.377		0.383		0.384		0.376		0.381		0.384	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Teacher-level model As a further advancement to the existing literature about reference-group effects on teachers' evaluations, we test whether our two indicators for teachers' frames of reference (which are dummy variables for teachers' more individualist and their more relational grade concepts, respectively) significantly interact with class-level predictors. Once again, we present separate models for class-level intelligence and class-level grades. We begin with the minimal *BFLPE* model including only class-level intelligence and class-level GPA (as well as the corresponding student-level predictor), respectively, to which we subsequently add each indicator of *TFR* plus its interaction with class-level achievement (tables 4 and 5). In several additional models, we add cross-level interactions between class- and student-level achievement as well as our student-level covariates (appendix, tables C to F).

As one can see, neither is one of the *TFR* indicators significantly associated with teachers' evaluations, nor does it significantly interact with class-level intelligence or GPA. Again, we test in terms of marginal effects and corresponding conditional z-values whether an insignificant interaction term was associated with significance differences in subgroups. Although these analyses also reveal that there are no differences in significance for both class-level intelligence and GPA by different categories of our two indicators of *TFR*, two observations appear to be worth mentioning:

First, apart from significance issues, students' effect of class-level intelligence on the probability of obtaining a positive teacher evaluation is slightly *weaker* for teachers who hold a more *individualist* grading concept, and slightly *stronger* for teachers who hold a more *relational* grading concept. Contrarily, the effect of class-average GPA is slightly *stronger* for teachers who hold a more *individualist* grading concept and slightly *weaker* for teachers who hold a more *relational* grading concept. To be sure, we cannot reject the null hypothesis that this finding is merely due to random error, which is why we are not able to maintain our hypothesis H_5.

Second, for one of the four interaction effects, the variance of the slopes by student-level subgroups appears to be a bit higher than for the three other one, namely for the effect of class-level GPA by teachers' relational grading concepts. These patterns are also illustrated by figure 4.

In several additional multilevel models, we repeat the analysis steps from the preceding tables in subsequently adding cross-level achievement interaction effects (*models 5a - 5h*) as well as student-level covariates (*models 6a - 6i*). Neither do we encounter remarkable differences compared to the models without *TFR* indicators, nor do the interactions of the latter terms with class-level achievement become significant or offer new insights regarding their marginal effects. Furthermore, no visible improvement in model fit could be gained. Therefore, these models are listed in the appendix (tables C – F).

Class-level model: SES In a last series of multilevel models, we tried to juxtapose potentially counterbalancing *BFLPE* and *RGE* patterns by including class-average SES in the model. In particular, we first estimated a reduced model containing only class-level SES and then added class-level achievement, their cross-level interaction terms with the corresponding student-level counterparts and additional student-level covariates. In

Table 4: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*), IST model

		Model 4a		Model 4b		Model 4c		Model 4d	
		Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value
	intercept	0.67***	3.54	0.67***	3.55	0.73***	3.88	0.73***	3.87
student level	ist	0.17***	18.01	0.17***	18.01	0.17***	18.01	0.17***	18
	gpa								
class level	IST	0.21***	5.02	0.21***	5.11	0.21***	5	0.2***	4.91
	GPA								
teacher level	TFR(ind)	0.16	1.07	0.15	1			-0.17	-0.76
	TFR(rel)					-0.22	-1.06		
cross-level interactions	IST*TFR(ind)			-0.03	-0.99				
	IST*TFR(rel)							0.04	0.73
	GPA*TFR(ind)								
	GPA*TFR(rel)								
ICC	student-level	0.654		0.654		0.654		0.654	
	class-level	0.106		0.106		0.107		0.107	
	teacher-level	0.094		0.094		0.093		0.093	
N	total	7627		7627		7627		7627	
	student	2774		2774		2774		2774	
	class	119		119		119		119	
	teacher	520		520		520		520	
fit measures	AIC	7174		7175		7174		7176	
	BIC	7223		7231		7223		7231	
	Nagelkerke's R^2	0.131		0.131		0.131		0.131	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table 5: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*), GPA model

		Model 4e		Model 4f		Model 4g		Model 4h	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
student level	intercept	0.52**	2.84	0.52**	2.84	0.58**	3.21	0.58**	3.2
	ist								
	gpa	7.35***	30.91	7.35***	30.92	7.35***	30.91	7.35***	30.92
class level	IST								
	GPA	3.27**	2.81	3.16**	2.67	3.31**	2.84	3.44**	2.89
teacher level	TFR(ind)	0.16	1.03	0.16	1.03				
	TFR(rel)					-0.23	-1.01	-0.18	-0.78
cross-level interactions	IST*TFR(ind)								
	IST*TFR(rel)								
	GPA*TFR(ind)			0.56	0.54				
	GPA*TFR(rel)							-0.92	-0.6
ICC	student-level	0.455		0.455		0.455		0.455	
	class-level	0.151		0.151		0.152		0.152	
	teacher-level	0.172		0.172		0.171		0.172	
	total	7626		7626		7626		7626	
N	student	2774		2774		2774		2774	
	class	119		119		119		119	
	teacher	520		520		520		520	
fit measures	AIC	5891		5893		5891		5893	
	BIC	5940		5948		5940		5948	
	Nagelkerke's R^2	0.352		0.352		0.352		0.352	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

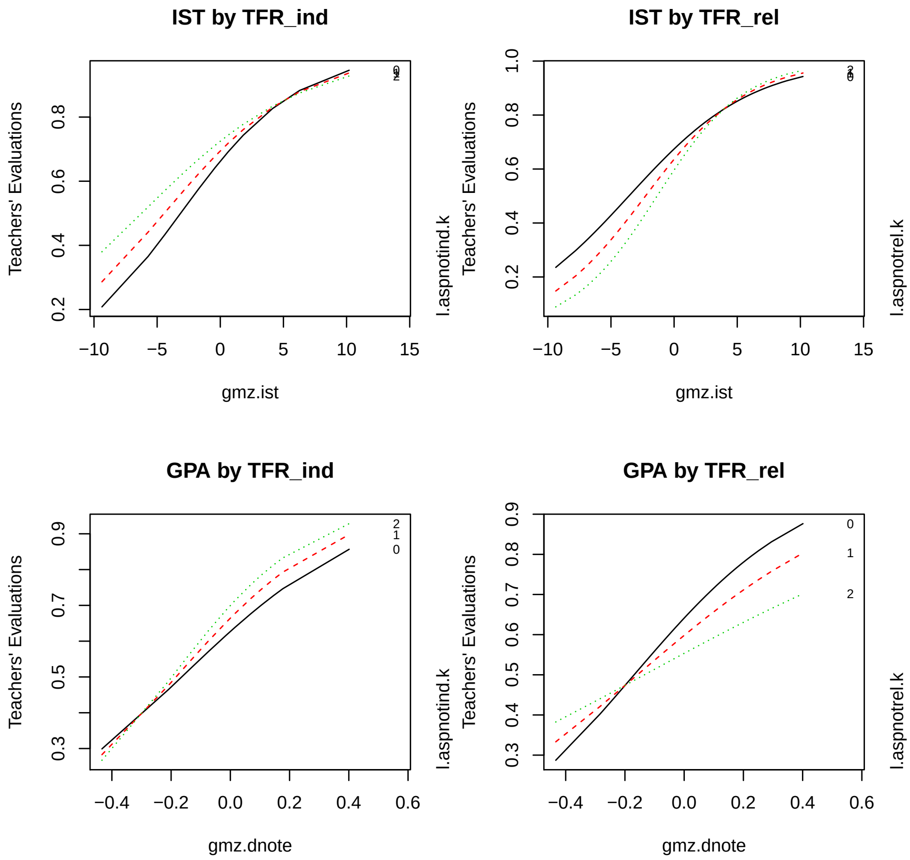


Figure 4: Marginal effects of cross-level interaction terms achievement * *TFR*.

model 7a, grand-mean centered class-average parental SES has a positive effect on the probability of obtaining a positive teacher evaluation. Also note that the model fit of this average-SES-only model is not worse than the fit of the pure achievement models of table 1. This could be a hint that a similar mechanism of reflected glory is at work when teachers 'upgrade' their evaluations in high-SES classes. In any case, our hypothesis H_2 is confirmed.

Adding class-level intelligence to the model (*model 7a*) does downsize the coefficient of class-level parental SES more than adding class-level GPA to the model (*model 7b*) – which could indicate that part of the class-average SES effect can be explained by a *primary* effect of social inequality (Boudon, 1974; Breen and Goldthorpe, 1997) that operates via intelligence differences between classrooms.[25] On the other hand, also when student-level intelligence goes into the model, class-average SES remains significant (to be precise, the effect even becomes a bit stronger), and it continues to exert its significant impact also when the models control for student-level GPA (which is – as in all other tables – by far the strongest predictor of teachers' evaluations). Concerning the cross-

[25] A mechanism that could account for this fact would be one of selective student-to-classroom sorting by both student intelligence and SES (Rothstein, 2009; Koedel and Betts, 2011).

level achievement interaction terms, neither do these coefficients alter the other model estimates, nor do we gain new evidence compared to the preceding tables.

In *models 8a-8h* (appendix, tables G and H), we add the remaining student-level covariates to the models from *models 7a-7g*. While the effect of class-average parental SES on teachers' evaluations remains significant in all of these models[26], regarding the student-level covariates, we do not find remarkable findings compared to the models without a class-level SES term. Whereas the cross-level interaction term between class-level intelligence and student-level intelligence remains insignificant, the cross-level interaction term between class-level and student level GPA is significant throughout the models except in *model 8h*. However, in the latter model, the marginal effects as well as the conditional significances for the `GPA*gpa` interaction term behave similar to the findings reported for table 2.

5 Discussion

Aim of study The objective of the paper at hand was to extend existing research about reference-group effects on teachers' evaluations. While previous studies based on Swiss and German data only analyzed primary school teachers' transition recommendations as an outcome (Trautwein and Baeriswyl, 2007; Tiedemann and Billmann-Mahecha, 2007; Gröhlich and Guill, 2009; Milek et al., 2009), we aimed to shed light on potential reference-group effects regarding the emergence of German 10^{th} class Gymnasium (i.e. highest track) teachers' evaluations of their students' prospective aptitude for academic studies at university. Dai and Rinn (2008) noted for classical *Big-Fish-Little-Pond Effect (BFLPE)* studies analyzing the effect of class-average achievement on students' self-concept that previous research has lacked a "specification of contexts where the BFLPE is more or less likely to occur" (p. 297). This is even more valid with regard to other outcomes such as teachers' recommendations or evaluations. Whereas in *BFLPE* research, at least moderator effects of student achievement (Marsh et al., 1995; Marsh and Rowe, 1996; Marsh et al., 2001; Marsh and Hau, 2003) and of *teachers' frames of reference* (*TFR*; Lüdtke et al., 2005) have been tested, similar analyses regarding teachers' evaluations as an outcome are lacking. Hence, we first tested whether potential reference-group effects in terms of class-average achievement on teachers' evaluations vary by students' achievement levels or *TFR*s in terms of teachers' more individualist or relational grading concepts, respectively.

Apart from these theoretical issues, our methodological contribution to the advancement of reference-group research was that we applied a more complex multilevel design wherein we regarded multiple teachers' evaluations of multiple students to be the lowest unit nested in both teacher and student contexts with school classes as an additional higher-level unit of analysis.

[26] Note that the effect of class-mean SES tends to increase when students' gender is controlled for. We will come back to this point in the discussion section.

Table 6: Multilevel analysis of teachers' evaluations: SES model

		Model 7a		*Model 7b*		*Model 7c*		*Model 7d*		*Model 7e*		*Model 7f*		*Model 7g*	
		Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value	Log-odds	z-value
	intercept	0.63***	3.7	0.64***	4.12	0.61***	3.77	0.72***	4.11	0.58***	3.43	0.72***	4.12	0.58***	3.47
student	ist							0.17***	17.94			0.17***	17.86		
level	gpa									7.36***	30.9			7.31***	30.81
class	IST			0.17***	4.73			0.17***	4.34			0.17***	4.31		
level	GPA					3.09**	2.89			2.71*	2.46			2.74*	2.49
	SES	1.56***	4.85	1.25***	4.11	1.45***	4.64	1.39***	4.05	1.32***	4.05	1.39***	4.05	1.33***	4.07
	IST*ist											0	1.27		
cross-level	GPA*gpa													3.14*	2.01
	student-level	0.656		0.677		0.666		0.67		0.469		0.668		0.465	
ICC	class-level	0.098		0.074		0.086		0.085		0.125		0.086		0.126	
	teacher-level	0.083		0.086		0.086		0.097		0.179		0.097		0.18	
	total	7968		7968		7968		7627		7626		7627		7626	
N	student	2903		2903		2903		2774		2774		2774		2774	
	class	119		119		119		119		119		119		119	
	teacher	520		520		520		520		520		520		520	
	AIC	7821		7804		7816		7816		5876		7160		5873	
fit	BIC	7856		7846		7858		7858		5925		7216		5929	
measures	Nagelkerke's R^2	0.005		0.008		0.006		0.134		0.355		0.134		0.355	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Interpretation of findings Based on the initial survey of the Cologne High School Panel (*CHiSP*; 1968/69), we found that in contrast to conventional *BFLPE* research with students' self-concept as an outcome (see research overviews by Trautwein and Lüdtke, 2005; Dai and Rinn, 2008; Dijkstra et al., 2008; and Marsh et al., 2008), but also in contrast to more recent reference-group effect studies of primary-school teachers' transition recommendations (Tiedemann and Billmann-Mahecha, 2007; Trautwein and Baeriswyl, 2007; Gröhlich and Guill, 2009), 10^{th} class Gymnasium teachers' evaluations of students' prospective aptitude for university studies (able vs. not able) are *positively* affected by both students' class-mean intelligence scores and class-mean grade point average (GPA). Hence, teachers are influenced by a kind of *Halo effect* (Thorndike, 1920) when evaluating their students; that is, regardless of a student's individual achievement, she gets a better evaluation in a higher-achieving class – and vice versa. This finding can be interpreted in a sense that teachers' beliefs about their students are positively affected by a higher-achieving learning environment – a result that contradicts our *BFLPE* or contrast hypothesis H_1 and supports our *RGE* hypothesis H_3 in its competing line of reasoning.

Although analyses have also shown that *given* a positive effect of class-level achievement on teachers' evaluations, student-level achievement is a far more important criterion for teachers in shaping their evaluations – indicated by remarkably higher Pseudo-R^2 values – we nonetheless owe the explanation *why* our results differ from reference-group effect research on primary-school teachers' transition recommendations that finds a *negative* class-average achievement effect. As a starting point, note that our analyses differs in three important respects from the latter studies: First, as already outlined in the hypotheses section, our data comprises a remarkably selective sample of Gymnasium (i.e. highest secondary track) students only, and possibly teachers are already primed beforehand in supposing their students to be relatively bright; and *if* they teach in a high-achievement class, the *Halo* effect is fostered by teachers' positive prejudices, and thus it is completely overshadowing negative contrast effects. Second, as opposed to more or less binding primary school teachers' transition recommendations, our outcome is a non-binding and even anonymous assessment without any direct consequences for students' educational life course. Perhaps, teachers more consciously consider the relative academic standing of a student compared to her classmates – which is an implication of the *BFLPE* hypothesis – when they know that their decision has a consequential (some would even say irreversible) impact on students' educational course of life. And third, our 10^{th} class teachers' evaluations are evidently measured at a later point in time than primary school teachers' transition recommendations. Although conventional *BFLPE* research has shown that the latter effect tends to be long-lasting (Marsh et al., 2000, 2001, 2007; Marsh and O'Mara, 2010), it is unknown whether the time of measurement of teachers' evaluations as an outcome makes a difference. Maybe as time goes by, teachers become more sure in their estimate of student-level ability, and so each student is judged more based on her own achievement than on her relative standing in class; and thus, the *BFLPE* is no longer powerful enough to overshadow its conceptual counterpart.

In any way, despite our positive class-level achievement effect that stands in contrast to existing studies up to now, results have also shown in terms of a much higher explained variance that student-level achievement is a much more important criterion for teachers in shaping their evaluations.

Regarding cross-level interaction effects, we found that the effect of class-average intelligence does not vary by student-level intelligence, but the effect of class-average GPA does become stronger for students with better marks – which stands in contrast to our expectations as outlined in interaction effect hypothesis H_4. Since the high achievers gain more from teachers' *Halo* effect than the low achievers, and teachers' expectations may affect students' later achievement (Rosenthal and Jacobson, 1968; Jussim and Harber, 2005; Becker, 2010; Birkelbach, 2011), this could result in cumulative educational inequality in sense of Merton's *Matthew effect* (Merton, 1968).

Although, as outlined in our hypotheses section, the alternative interpretation of this interaction term in that the effect of student-level GPA on teachers' evaluations varies by class-level GPA is also possible – which would mean that students' individual achievement becomes more distinct in high achievement contexts –, the marginal effects of this (and also of the `ist*IST`) interaction term show less variance compared to the specification in our models (cf. figure A, appendix).

Contrarily, we did not find evidence for reference-group effects to vary by teachers' grading concepts according to conventional significance thresholds. A graphical inspection revealed that there is an (admittedly insignificant) tendency in that the effect of class-level intelligence on the probability of obtaining a positive teacher evaluation is slightly *weaker* for teachers who hold a more *individualist* grading concept, and slightly *stronger* for teachers who hold a more *relational* grading concept; while the effect of class-average GPA is slightly *stronger* for teachers who hold a more *individualist* grading concept and slightly *weaker* for teachers who hold a more *relational* grading concept. Although apart from significance issues, we would not consider it to be implausible that for individualist teachers, class-average intelligence becomes a less important criterion in forming their evaluations, we would have difficulties to explain the latter case, and therefore simply concede that we are not able to maintain our underlying interaction effect hypothesis H_5.

In contrast to Marsh et al. (2000), we were not able to juxtapose potentially counterbalancing negative class-level achievement effects and positive class-level SES effects. Apart from the aforementioned positive class-average achievement effects we found, we additionally encountered a similarly positive class-average SES effect which remains significant throughout all models. This finding could be due to the fact that the students in our sample have been surveyed in 10^{th} class of Gymnasium, and according to relative risk aversion theory (Breen and Goldthorpe, 1997), the higher a student went on the educational ladder, the less likely social class differences will be able to explain differences in achievement since selection effects based on social class have already taken place; and only the most able (or motivated) lower-class students remain in the higher school tracks. On the other hand, the effect of class-level SES becomes even stronger when additionally to class-average intelligence, also student-level intelligence is controlled for; and a similar trend can be observed when students' gender is added to the model. Apparently, if differences within classes with regard to these two student characteristics are ruled out, the 'true' between-class SES effect is unmasked.

Another issue is worthy of discussion, too: In contrast to both Gröhlich and Guill (2009) and Milek et al. (2009), we did not find contextual-level achievement effects to be mediated by students' individual achievement. But these two studies analyzed reference-group effects on explicit (and more binding) primary-school teachers' transition recommendations, and in our case of implicit and non-binding 10^{th} class teachers' evaluations regarding their students' prospective aptitude for academic studies at university, teachers might be more error-prone in terms of being influenced by contextual-level effects. Another possible explanation would be that only negative reference-group effects are mediated by student-level achievement but not positive ones, since *Halo* effects possibly evoke other (e.g. more motivation-related) teacher expectancies than cannot be explained by student grades.

In sum, our result of positive class-level achievement effects on teachers' evaluations indicate to reject our *BFLPE* hypothesis H_1 and to maintain its competitor, the *RGE* hypothesis in terms of a *Halo* effect (H_3). Similarly, we also found a positive effect of class-average socioeconomic status (SES) – which confirmed our hypothesis H_2, – but we were not able to juxtapose a potentially still present *BFLPE* from our *RGE* when controlling for both class-level achievement and SES. Regarding interaction effects, we found that the effect of class-level GPA gets stronger with increasing student-level GPA, which contradicts our hypothesis H_4. Moreover, we found quite puzzling results for the variance of class-level IST and GPA by teachers' frames of reference; but as both marginal effects were not significant, we simply reject our underlying hypothesis H_5.

Advice for further studies Taken together, our findings lead to a couple of new questions that could be answered in further studies. First, taking Dai and Rinn's (2008) remark for serious, further tests of both student and teacher-level moderators are highly recommended. On the one hand, more precise evidence on the mechanisms behind the variance of reference-group effects on teachers' evaluations by student-level achievement has to be collected. Remarkably, this effect only concerned school grades – definitely comprising an evaluative component as well. On the other hand, we noted a quite puzzling (though insignificant) pattern regarding the interaction between class-level achievement and teachers' frames of reference. Further studies should try to operationalize these frames by survey items that are specified more explicitly in order to collect richer data on teachers' grading concepts that allows to arrive at conclusions fulfilling conventional significance thresholds.

Since we observed that the effect of class-average achievement is not completely insensitive against controls for students' gender[27], prospective research on reference-group effects regarding teachers' evaluations should also include gender composition at the classroom level in the analysis. Thijs et al. (2010) have shown that the gender compo-

[27]While based on more recent data, Gröhlich and Guill (2009) found male students to obtain worse evaluations by their teachers (in terms of transition recommendations), in our study, it was just vice versa (controlled for achievement, respectively). A possible explanation could be that teachers' frames of gender-related educational inequality changed over time.

sition of the classroom is a promising moderator of conventional *BFLPE* research with students' self-concept as an outcome. Although, at least in Germany, two recent studies observed students' school grades to be influenced neither by same-gender effects nor by teachers' gender composition (Neugebauer et al., 2011; Helbig, 2010), an analysis of how teachers' evaluations might be affected by particular combinations of a student's gender, a teacher's gender, student-gender composition effects in the classroom and maybe even teacher-gender composition effects at school is still missing.

In methodological terms, this question could also be answered in the framework of a cross-classified hierarchical model. While our study simplified analyses by looking at the dichotomy of students with a positive vs. those with a negative evaluation, other analyses with an ordinal factor variable as an outcome could apply an ordered categorical multilevel model (Johnson, 1996, 1997; Gelman and Hill, 2007, p. 331f.) in order to use all available information. However, note that these models would have to be employed in a Bayesian framework.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716–723.

Alexander K and Eckland BK (1975) Contextual effects in the high school attainment process. *American Sociological Review* 40(3): 402–416.

Alexander KL; Entwisle DR and Thompson MS (1987) School performance, status relations, and the structure of sentiment: bringing the teacher back in. *American Sociological Review* 52(5): 665–682.

Alwin DF and Otto LB (1977) High school context effects on aspirations. *Sociology of Education* 50(4): 259–273.

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

Arnold KH; Bos W; Richert P et al. (2007) Schullaufbahnpräferenzen am Ende der vierten Klassenstufe. In: Bos W; Hornberg S; Arnold KH et al. (eds.) *IGLU 2006 - Schullaufbahnempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe*, 271–297. Münster / New York: Waxmann.

Bates D and Maechler M (2009) lme4 - Linear mixed-effects models using S4 classes. R package version 0.999375-31.
URL: *http://cran.r-project.org/web/packages/lme4/index.html*

Baumert J; Brunner M; Lüdtke O et al. (2007) Was messen internationale Schulleistungsstudien? Resultate kumulativer Wissenserwerbsprozesse. *Psychologische Rundschau* 58(2): 118–128.

Becker D (2010) The impact of teachers' expectations on students' educational opportunities in the life course. *Cologne Graduate School Working Paper Series* 1(7). **URL:** *http://econpapers.repec.org/paper/cgrcgsser/01-07.htm*

Becker D and Birkelbach K (2010) Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond- oder Reflected-Glory-Effekt? Eine Mehrebenen-Analyse von Lehrerurteilen. In: Beckers T; Birkelbach K; Hagenah J et al. (eds.) *Komparative empirische Sozialforschung*, 113–141. Wiesbaden: VS-Verlag.

Becker D and Birkelbach K (2011) Teachers' evaluations and the definition of the social situation in the classroom. *Cologne Graduate School Working Paper Series* 2(4). **URL:** *http://ideas.repec.org/p/cgr/cgsser/02-04.html*

Birkelbach K (2011) Teacher evaluations over the life course: valid prognosis or self-fulfilling prophecy? In: Mica A; Peisert A and Winczorek J (eds.) *Sociology and the Unintended - Robert Merton Revisited*, 133–153. Frankfurt a.M.: Peter Lang.

Bos W and Pietsch M (2004) *Erste Ergebnisse aus KESS 4. Kurzbericht.* Hamburg: Freie und Hansestadt Hamburg.

Bos W; Voss A; Lankes EM et al. (2004) Schullaufbahnempfehlungen von Lehrkräften für Kinder am Ende der vierten Jahrgangsstufe. In: Bos W; Lankes EM; Prenzel M et al. (eds.) *IGLU. Einige Länder der Bundesrepublik Deutschland im nationalen und internationalen Vergleich*, 191–228. Münster: Wachsmann.

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

Bourdieu P (1986) *Distinction: A Social Critique of the Judgement of Taste.* London: Routledge.

Bourdieu P and Passeron JC (1990) *Reproduction in Education, Society and Culture.* London: Sage Publications Ltd.

Brambor T; Clark WR and Golder M (2006) Understanding interaction models: improving empirical analyses. *Political Analysis* 14(1): 63–82.

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.

Brophy JE and Good TL (1974) *Teacher - Student Relationships. Causes and Consequences.* New York: Holt, Rhinehart and Winston.

Bryk AS and Raudenbush SW (1992) *Hierarchical Linear Models.* Thousand Oaks, CA: Sage Publications.

Burnham KP and Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach.* 2nd edn. New York: Springer.

Burnham KP and Anderson DR (2004) Multimodel inference. Understanding AIC and BIC in model selection. *Sociological Methods & Research* 33(2): 261–304.

Burton NW and Ramist L (2001) *Predicting Success in College: SAT Studies of Classes Graduating Since 1980.* New York: College Board Research Report No. 2001-2.

Camara W (1998) *High School Grading Policies.* Tech. Rep. RN-04, College Board, New York.

Camara W and Echternacht G (2000) *The SAT I and high school grades: Utility in predicting success in college.* New York: The College Board.

Camara W; Kimmel E; Scheuneman J et al. (2003) *Whose Grades Are Inflated?* Tech. Rep. RN-10, College Board, New York.

Chanal J; Marsh HW; Sarrazin P et al. (2005) Big-fish-little-pond effects on gymnastics self-concept: social comparison processes in a physical setting. *Journal of Sport and Exercise Psychology* 27(1): 53–70.

Cialdini RB; Borden RJ; Thorne A et al. (1976) Basking in reflected glory: three (football) field studies. *Journal of Personality and Social Psychology* 34(3): 366–375.

Coleman JM and Fults BA (1985) Special-class placement, level of intelligence, and the self-concepts of gifted children: a social comparison perspective. *Remedial and Special Education* 6(1): 8–10.

Dai DY and Rinn AN (2008) The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review* 20(3): 283–317.

Davis JA (1966) The campus as a frog pond: an application of the theory of relative deprivation to career decisions of college men. *The American Journal of Sociology* 72(1): 17–31.

DeLeeuw J (1992) Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle. In: Kotz S and Johnson NL (eds.) *Breakthroughs in Statistics, vol. 1*, 559–609. London: Springer.

Dijkstra P; Kuyper H; van der Werf G et al. (2008) Social comparison in the classroom: a review. *Review of Educational Research* 78(4): 828–879.

Downey DB and Pribesh S (2004) When race matters: teachers' evaluations of students' classroom behavior. *Sociology of Education* 77(4): 267–282.

Erikson RS; Goldthorpe JH; Jackson M et al. (2005) On class differentials in educational attainment. *Proceedings of the National Academy of Sciences of the United States of America* 102(27): 9730–9733.

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

Farkas G; Grobe RP; Sheehan D et al. (1990) Cultural resources and school success: gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review* 55(1): 127–142.

Friedrich RJ (1982) In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science* 26(4): 797–833.

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

Goldstein H (1995) *Multilevel Statistical Methods.* London: Hodder Arnold.

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.

Good TL and Brophy JE (2003) *Looking in Classrooms.* New York: Pearson Education.

Gröhlich C and Guill K (2009) How stable are reference group effects of secondary school track recommendations? *Journal for Educational Research Online/Journal für Bildungsforschung Online* 1(1): 154–171.

Helbig M (2010) Sind Lehrerinnen für den geringeren Schulerfolg von Jungen verantwortlich? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 62(1): 93–111.

Hox JJ (2010) *Multilevel Analysis - Techniques and Applications.* 2nd edn. New York and Hove: Routledge.

Ingenkamp K (1977) *Schüler- und Lehrerbeurteilung. Empirische Untersuchungen zur pädagogischen Diagnostik.* Weinheim: Beltz Verlag.

Ingenkamp KH (1993) Der Prognosewert von Zensuren, Lehrergutachten, Aufnahmeprüfungen und Tests während der Grundschulzeit für den Sekundarschulerfolg. In: Olechowski R (ed.) *Frühe Schulische Auslese*, 68–85. Frankfurt: Lang.

Johnson VE (1996) On Bayesian analysis of multirater ordinal data: an application to automated essay grading. *Journal of the American Statistical Association* 91(433): 42–51.

Johnson VE (1997) An alternative to traditional GPA for evaluating student performance. *Statistical Science* 12(4): 251–269.

Jürges H and Schneider K (2006) Age at school entry and teachers' recommendations for secondary school track choice in Germany. *Paper presented at the 20th annual ESPE conference, 22 - 24 June, 2006, Verona* .

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.

Koedel C and Betts JR (2011) Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy* 6(1): 18–42.

Kozlowski SW and Kirsch MP (1987) The systematic distortion hypothesis, halo, and accuracy: an individual-level analysis. *Journal of Applied Psychology* 72(2): 252–261.

Kreft IGG; De Leeuw J and Aiken LS (1995) The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research* 30(1): 1–21.

Kristen C (2006) Ethnische Diskriminierung in der Grundschule? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58(1): 79–97.

Kuncel NR; Hezlett SA and Ones DS (2001) A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: implications for graduate student selection and performance. *Psychological Bulletin* 127(1): 162–181.

Lehmann RH; Peek R and GänsfußR (1997) Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996.

Long JS (1997) *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

Lüdtke O; Köller O; Marsh HW et al. (2005) Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology* 30(3): 263–285.

Maas CJM and Hox JJ (2005) Sufficient sample sizes for multilevel modeling. *Methodology* 1(3): 86–92.

Maaz K; Hausen C; Mcelvany N et al. (2006) Stichwort: Übergänge im Bildungssystem. *Zeitschrift für Erziehungswissenschaft* 9(3): 299–327.

Marsh HW (1987) The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology* 79(3): 280–295.

Marsh HW (1991) The failure of high ability high schools to deliver academic benefits: the importance of academic self-concept and educational aspirations. *American Educational Research Journal* 28(2): 445–480.

Marsh HW; Chessor D; Craven RG et al. (1995) The effects of gifted and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal* 32(2): 285–319.

Marsh HW; Chit-Kwong K and Kit-Tai H (2000) Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology* 78(2): 337–349.

Marsh HW and Craven RG (2006) Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science* 1(2): 133–163.

Marsh HW and Hau KT (2003) Big-fish-little-pond effect on academic self-concept: a cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist* 58(5): 364–376.

Marsh HW; Köller O and Baumert J (2001) Reunification of East and West German school systems: longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal* 38(2): 321–350.

Marsh HW and O'Mara AJ (2010) Long-term total negative effects of school-average ability on diverse educational outcomes. *Zeitschrift für Pädagogische Psychologie* 24(1): 51–72.

Marsh HW and Parker JW (1984) Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology* 47(1): 213–231.

Marsh HW and Rowe KJ (1996) The negative effects of school-average ability on academic self-concepts: an application of multilevel modelling. *Australian Journal of Education* 40(1): 65–87.

Marsh HW; Seaton M; Trautwein U et al. (2008) The big-fish–little-pond-effect stands up to critical scrutiny: implications for theory, methodology, and future research. *Educational Psychology Review* 20(3): 319–350.

Marsh HW; Trautwein U; Lüdtke O et al. (2005) Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development* 76(2): 397–416.

Marsh HW; Trautwein U; Lüdtke O et al. (2007) The big-fish-little-pond effect: persistent negative effects of selective high schools on self-concept after graduation. *American Educational Research Journal* 44(3): 631–669.

McFarland C and Buehler R (1995) Collective self-esteem as a moderator of the frog-pond effect in reactions to performance feedback. *Journal of Personality and Social Psychology* 68(6): 1055–1070.

Mechtenberg L (2009) Cheap talk in the classroom: How biased grading at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies* 76(4): 1431–1459.

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

Merton RK (1968) The Matthew effect in science. *Science* 159(3810): 56–63.

Meulemann H (1979) *Soziale Herkunft und Schullaufbahn. Arbeitsbuch zur sozialwissenschaftlichen Methodenlehrer.* Frankfurt / New York: Campus Verlag.

Meyer JW (1970) High school effects on college intentions. *The American Journal of Sociology* 76(1): 59–70.

Milek A; Lüdtke O; Trautwein U et al. (2009) Wie konsistent sind Referenzgruppeneffekte bei der Vergabe von Schulformempfehlungen? Bundeslandspezifische Analysen mit Daten der IGLU-Studie. *Zeitschrift für Erziehungswissenschaft* 12(12): 282–301.

Morgan R (1989) *Analyses of the Predictive Validity of the SAT and High School Grades From 1976 to 1985.* New York: College Board Research Report No. 89-7.

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

Muthen BO and Satorra A (1995) Complex sample data in structural equation modeling. *Sociological Methodology* 25: 267–316.

Neugebauer M; Helbig M and Landmann A (2011) Unmasking the myth of the same-sex teacher advantage. *European Sociological Review* 27(5): 669–689.

Pallas AM; Entwisle DR; Alexander KL et al. (1994) Ability-group effects: instructional, social, or institutional? *Sociology of Education* 67(1): 27–46.

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.

Plucker JA; Robinson NM; Greenspon TS et al. (2004) It's not how the pond makes you feel, but rather how high you can jump. *American Psychologist* 59(4): 268–269.

Raftery AE (1995) Bayesian model selection in social research. *Sociological Methodology* 25: 111–164.

Raudenbush SW; Yang ML and Yosef M (2000) Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* 9(1): 141–157.

Reuman DA (1989) How social comparison mediates the relation between ability-grouping practices and students' achievement expectancies in mathematics. *Journal of Educational Psychology* 81(2): 178–189.

Rheinberg F (1980) *Leistungsbewertung und Lernmotivation.* Göttingen: Hogrefe.

Rindermann H and Heller KA (2005) The benefit of gifted classes and talent schools for developing students' competences and enhancing academic self-concept. *Zeitschrift für Pädagogische Psychologie* 19(3): 133–136.

Robbins SB; Lauver K; Le H et al. (2004) Do psychosocial and study skill factors predict college outcomes? a meta-analysis. *Psychological Bulletin* 130(2): 261–288.

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

Rothstein J (2009) Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy* 4(4): 537–571.

Schwarz G (1978) Estimating the dimension of a model. *The Annals of Statistics* 6(2): 461–464.

Seaton M; Marsh HW and Craven RG (2009) Earning its place as a pan-human theory: universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology* 101(2): 403–419.

Seaton M; Marsh HW and Craven RG (2010) Big-fish-little-pond effect: generalizability and moderation - two sides of the same coin. *American Educational Research Journal* 47(2): 390–433.

Snijders TAB and Bosker RJ (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage Publications Ltd.

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.

Tabachnick BG and Fidell LS (2007) *Using Multivariate Statistics.* 5 edn. Boston, Mass: Pearson Education.

Thijs J; Verkuyten M and Helmond P (2010) A further examination of the big-fish-little-pond effect. *Sociology of Education* 83(4): 333–345.

Thorndike EL (1920) A constant error in psychological ratings. *Journal of Applied Psychology* 4(1): 25–29.

Tiedemann J and Billmann-Mahecha E (2007) Zum Einfluss von Migration und Schulklassenzugehörigkeit auf die Übergangsempfehlung für die Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft* 10(1): 108–120.

Trautwein U and Baeriswyl F (2007) Wenn leistungsstarke Klassenkameraden ein Nachteil sind. *Zeitschrift für Pädagogische Psychologie* 21(2): 119–133.

Trautwein U; Gerlach E and Lüdtke O (2008) Athletic classmates, physical self-concept, and free-time physical activity: a longitudinal study of frame of reference effects. *Journal of Educational Psychology* 100(4): 988–1001.

Trautwein U and Lüdtke O (2005) The big-fish-little-pond effect. Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie* 19(3): 137–140.

Trautwein U; Lüdtke O; Marsh HW et al. (2006) Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology* 98(4): 788–806.

Trautwein U; Lüdtke O; Marsh HW et al. (2009) Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology* 101(4): 853–866.

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* 57(2): 307–333.

Zeidner M and Schleyer EJ (1998) The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology* 24(4): 305–329.

6 Appendix

Table A: Descriptive statistics of independent variables

	count	mean	sd	min	max
ist (student-level intelligence, group-mean centered)	16819	0.01	10.45	-35.59	37.17
gpa (student-level GPA, group-mean centered)	16820	0.00	0.45	-1.35	1.73
ses (student-level SES, group-mean centered)	16833	0.00	1.19	-3.52	3.19
aspir (student-level aspirations)	16796	0.68	0.47	0.00	1.00
sex (student-level gender)	17689	0.54	0.50	0.00	1.00
IST (class-level intelligence, grand-mean centered)	17689	-0.12	4.45	-9.37	10.20
GPA (class-level GPA, grand-mean centered)	17689	-0.00	0.15	-0.43	0.40
SES (class-level SES, grand-mean centered)	17689	0.01	0.52	-1.27	1.46
TFR(ind) (Teachers' frame of reference: individualist)	17626	0.22	0.59	0.00	2.00
TFR(rel) (Teachers' frame of reference: relational)	17626	0.10	0.42	0.00	2.00

Table B: Correlation matrix of independent variables

	ist	gpa	ses	aspir	sex	IST	GPA	SES	TFR(ind)	TFR(rel)
ist	1									
gpa	.247***	1								
ses	-.00922	.0791***	1							
aspir	.0631***	.140***	.0979***	1						
sex	.0171*	-.0543***	-.00686	.131***	1					
IST	.00168	-.00112	.00382	.138***	.386***	1				
GPA	.000190	-.00160	.00303	.0279***	-.199***	.0343***	1			
SES	.000817	-.00153	.00546	.128***	-.102***	.218***	.145***	1		
TFR(ind)	-.000874	.000813	-.000878	.00536	.00459	-.0128	.0218**	.0673***	1	
TFR(rel)	-.000500	.000522	-.00259	-.0123	-.0381***	-.0207**	.0338***	-.0247**	-.0268***	1

* $p < .05$, ** $p < .01$, *** $p < .001$

Table C: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*) + cross-level achievement interaction terms, IST model

		Model 5a		*Model 5b*		*Model 5c*		*Model 5d*	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
	intercept	0.56**	3.19	0.56**	3.2	0.61***	3.52	0.62***	3.54
student	ist	0.06***	8.49	0.06***	8.49	0.06***	8.49	0.06***	8.49
level	gpa	6.96***	29.73	6.96***	29.73	6.96***	29.74	6.95***	29.74
class	IST	0.18***	4.67	0.18***	4.68	0.17***	4.51	0.17***	4.52
level	GPA								
teacher	TFR(ind)	0.15	0.93	0.15	0.93				
level	TFR(rel)					-0.16	-0.67	-0.16	-0.67
	IST*ist			0	0.67			0	0.67
	GPA*gpa								
cross-level	IST*TFR(ind)	-0.03	-0.88	-0.03	-0.88				
interactions	IST*TFR(rel)					0.02	0.38	0.02	0.38
	GPA*TFR(ind)								
	GPA*TFR(rel)								
	student-level	0.449		0.449		0.449		0.449	
ICC	class-level	0.139		0.139		0.141		0.141	
	teacher-level	0.177		0.177		0.176		0.176	
	total	7602		7602		7602		7602	
N	student	2764		2764		2764		2764	
	class	119		119		119		119	
	teacher	520		520		520		520	
	AIC	5771		5773		5772		5774	
fit	BIC	5834		5842		5834		5843	
measures	Nagelkerke's R^2	0.372		0.372		0.372		0.372	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table D: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*) + cross-level achievement interaction terms, GPA model

		Model 5e		*Model 5f*		*Model 5g*		*Model 5h*	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
student level	intercept	0.52**	2.85	0.53**	2.88	0.58**	3.21	0.59**	3.25
	ist	0.06***	8.47	0.06***	8.47	0.06***	8.46	0.06***	8.46
	gpa	6.95***	29.73	6.91***	29.63	6.95***	29.73	6.91***	29.63
class level	IST								
	GPA	3.09**	2.61	3.13**	2.64	3.38**	2.84	3.43**	2.89
teacher level	TFR(ind)	0.16	1.03	0.16	1.03				
	TFR(rel)					-0.19	-0.83	-0.19	-0.81
cross-level interactions	IST*ist								
	GPA*gpa			2.98	1.94			2.98	1.94
	IST*TFR(ind)								
	IST*TFR(rel)								
	GPA*TFR(ind)	0.64	0.62	0.69	0.67				
	GPA*TFR(rel)					-0.86	-0.57	-0.96	-0.63
ICC	student-level	0.438		0.435		0.437		0.435	
	class-level	0.159		0.16		0.16		0.16	
	teacher-level	0.174		0.175		0.174		0.175	
N	total	7602		7602		7602		7602	
	student	2764		2764		2764		2764	
	class	119		119		119		119	
	teacher	520		520		520		520	
fit measures	AIC	5784		5781		5784		5781	
	BIC	5847		5851		5846		5851	
	Nagelkerke's R^2	0.37		0.37		0.37		0.37	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table E: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*) + student-level covariates, IST model

		Model 6a		*Model 6b*		*Model 6c*		*Model 6d*		*Model 6e*		*Model 6f*	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
student level	intercept	0.56**	3.18	0.12	0.56	-0.22	-0.9	0.61***	3.53	0.17	0.83	-0.16	-0.67
	ist	0.07***	8.66	0.06***	8.46	0.06***	8.36	0.07***	8.66	0.06***	8.46	0.06***	8.36
	gpa	6.86***	29.55	6.77***	29.1	6.83***	29.03	6.86***	29.56	6.77***	29.1	6.82***	29.03
	ses	0.28***	4.45	0.25***	3.98	0.25***	4.02	0.28***	4.45	0.25***	3.97	0.25***	4.01
	aspir			0.65***	3.89	0.62***	3.73			0.64***	3.89	0.62***	3.73
	sex					0.64*	2.56					0.63*	2.55
class level	IST	0.18***	4.68	0.17***	4.49	0.15***	3.85	0.17***	4.52	0.16***	4.35	0.14***	3.7
	GPA												
teacher level	TFR(ind)	0.15	0.96	0.15	0.94	0.14	0.93						
	TFR(rel)							-0.18	-0.77	-0.18	-0.74	-0.17	-0.7
cross-level interactions	IST*ist	0	0.73	0	0.7	0	0.77	0	0.73	0	0.7	0	0.78
	GPA*gpa												
	IST*TFR(ind)	-0.03	-0.85	-0.02	-0.77	-0.02	-0.76	0.02	0.38	0.02	0.35	0.02	0.34
	IST*TFR(rel)												
	GPA*TFR(ind)												
	GPA*TFR(rel)												
ICC	student-level	0.443		0.446		0.451		0.442		0.445		0.45	
	class-level	0.142		0.137		0.129		0.144		0.139		0.131	
	teacher-level	0.178		0.178		0.179		0.176		0.176		0.177	
N	total	7588		7552		7552		7588		7552		7552	
	student-level	2759		2747		2747		2759		2747		2747	
	class-level	119		119		119		119		119		119	
	teacher-level	520		519		519		520		519		519	
fit measures	AIC	5744		5711		5706		5745		5712		5707	
	BIC	5821		5794		5797		5821		5795		5797	
	Nagelkerke's R^2	0.377		0.383		0.384		0.377		0.383		0.384	

All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table F: Multilevel analysis of teachers' evaluations: Teachers' frames of reference (*TFR*) + student-level covariates, GPA model

		Model 6g Log-odds	*z*-value	*Model 6h* Log-odds	*z*-value	*Model 6i* Log-odds	*z*-value	*Model 6j* Log-odds	*z*-value	*Model 6k* Log-odds	*z*-value	*Model 6l* Log-odds	*z*-value
student level	intercept	0.52**	2.86	0.07	0.33	-0.43	-1.82	0.59**	3.24	0.13	0.64	-0.37	-1.57
	ist	0.06***	8.64	0.06***	8.44	0.06***	8.28	0.06***	8.64	0.06***	8.43	0.06***	8.28
	gpa	6.82***	29.44	6.72***	28.98	6.82***	28.94	6.81***	29.44	6.72***	28.98	6.81***	28.94
	ses	0.28***	4.51	0.25***	4.01	0.26***	4.06	0.28***	4.5	0.25***	4	0.26***	4.05
	aspir			0.66***	4.02	0.62***	3.77			0.66***	4.02	0.62***	3.76
	sex					0.96***	3.98					0.96***	3.96
class level	IST												
	GPA	3.12**	2.63	3.11**	2.68	3.7***	3.4	3.39**	2.84	3.36**	2.88	3.94***	3.6
teacher level	TFR(ind)	0.16	1.05	0.16	1.03	0.15	0.98						
	TFR(rel)							-0.22	-0.96	-0.21	-0.92	-0.19	-0.84
cross-level interactions	IST*ist												
	GPA*gpa	3.13*	2.06	2.94	1.93	2.83	1.85	3.12*	2.05	2.93	1.92	2.82	1.85
	IST*TFR(ind)												
	IST*TFR(rel)												
	GPA*TFR(ind)	0.68	0.67	0.56	0.55	0.57	0.56						
	GPA*TFR(rel)							-0.68	-0.45	-0.7	-0.46	-0.67	-0.45
ICC	student-level	0.429		0.434		0.451		0.428		0.433		0.45	
	class-level	0.163		0.154		0.124		0.164		0.155		0.125	
	teacher-level	0.175		0.176		0.182		0.175		0.176		0.181	
N	total	7588		7552		7552		7588		7552		7552	
	student-level	2759		2747		2747		2759		2747		2747	
	class-level	119		119		119		119		119		119	
	teacher-level	520		519		519		520		519		519	
fit measures	AIC	5752		5718		5705		5752		5718		5705	
	BIC	5829		5802		5795		5829		5801		5796	
	Nagelkerke's R^2	0.376		0.383		0.384		0.376		0.383		0.384	

All coefficients are unstandardized log odds. Significance values: * (p < .05); ** (p < .01); *** (p < .001).

Table G: Multilevel analysis of teachers' evaluations: SES model with student-level covariates (IST model)

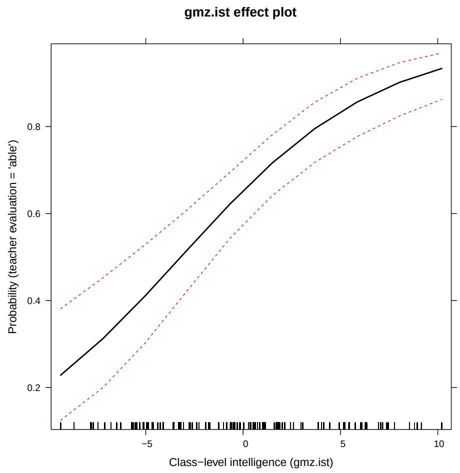
		Model 8a		*Model 8b*		*Model 8c*		*Model 8d*	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
student level	intercept	0.61***	3.74	0.61***	3.71	0.18	0.91	-0.23	-0.99
	ist	0.06***	8.48	0.07***	8.65	0.06***	8.45	0.06***	8.33
	gpa	6.96***	29.74	6.86***	29.55	6.78***	29.1	6.85***	29.05
	ses			0.28***	4.47	0.25***	4	0.26***	4.06
	aspir					0.62***	3.75	0.59***	3.52
	sex							0.79**	3.21
class level	IST	0.15***	3.99	0.15***	4	0.14***	3.84	0.11**	2.97
	GPA								
	SES	1.14***	3.59	1.15***	3.59	1.09***	3.42	1.22***	3.96
interaction terms	IST*ist	0	0.68	0	0.74	0	0.71	0	0.81
	GPA*gpa								
ICC	student-level	0.459		0.453		0.454		0.463	
	class-level	0.118		0.121		0.118		0.104	
	teacher-level	0.184		0.184		0.184		0.186	
N	total	7602		7588		7552		7552	
	student-level	2764		2759		2747		2747	
	class-level	119		119		119		119	
	teacher-level	520		520		519		519	
fit-measures	AIC	5760		5731		5699		5691	
	BIC	5822		5801		5775		5774	
	Nagelkerke's R^2	0.373		0.379		0.384		0.386	

All coefficients are unstandardized log odds. Significance values: * (p < .05); ** (p < .01); *** (p < .001).

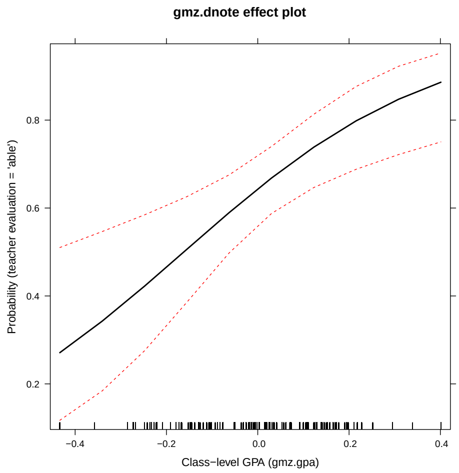
Table H: Multilevel analysis of teachers' evaluations: SES model with student-level covariates (GPA model)

		Model 8e		*Model 8f*		*Model 8g*		*Model 8h*	
		Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value	Log-odds	*z*-value
	intercept	0.59***	3.48	0.59***	3.46	0.15	0.74	-0.41	-1.86
student	ist	0.06***	8.45	0.06***	8.63	0.06***	8.43	0.06***	8.25
level	gpa	6.91***	29.62	6.82***	29.43	6.73***	28.97	6.83***	28.96
	ses			0.28***	4.52	0.25***	4.04	0.26***	4.1
	aspir					0.63***	3.84	0.58***	3.51
	sex							1.09***	4.64
class	IST								
level	GPA	2.71*	2.45	2.69*	2.44	2.7*	2.48	3.35***	3.38
	SES	1.32***	4.04	1.32***	4.04	1.24***	3.86	1.31***	4.48
interaction	IST*ist								
terms	GPA*gpa	3.01*	1.97	3.16*	2.08	2.99*	1.96	2.88	1.89
	student-level	0.448		0.441		0.445		0.466	
ICC	class-level	0.133		0.136		0.131		0.092	
	teacher-level	0.182		0.183		0.183		0.19	
	total	7602		7588		7552		7552	
N	student-level	2764		2759		2747		2747	
	class-level	119		119		119		119	
	teacher-level	520		520		519		519	
	AIC	5765		5736		5703		5685	
fit	BIC	5827		5805		5779		5769	
measures	Nagelkerke's R^2	0.373		0.377		0.384		0.387	

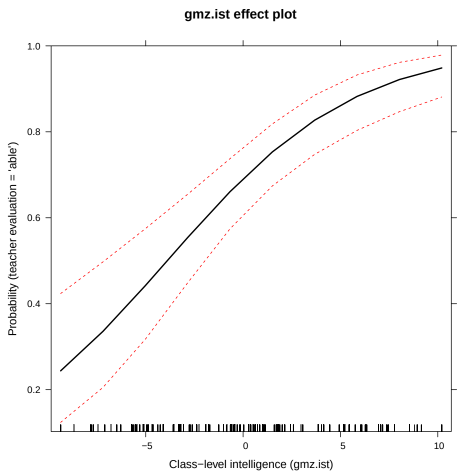
All coefficients are unstandardized log odds. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).



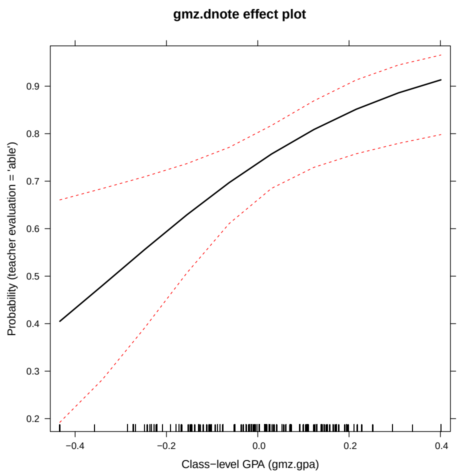
(a) Model 1a.



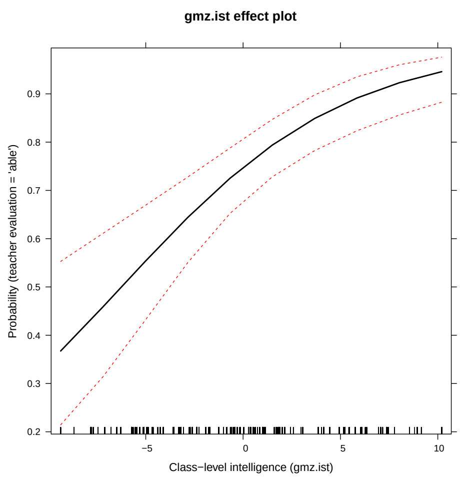
(b) Model 1b.



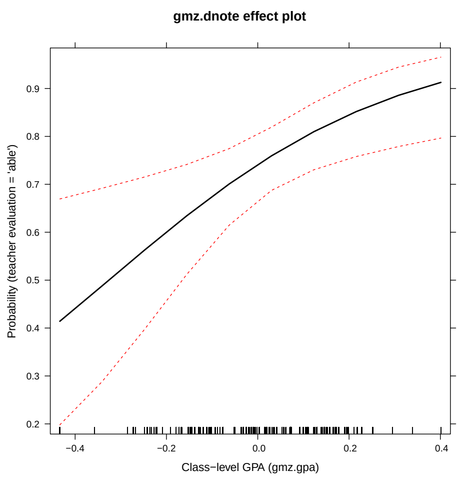
(c) Model 1c.



(d) Model 1d.

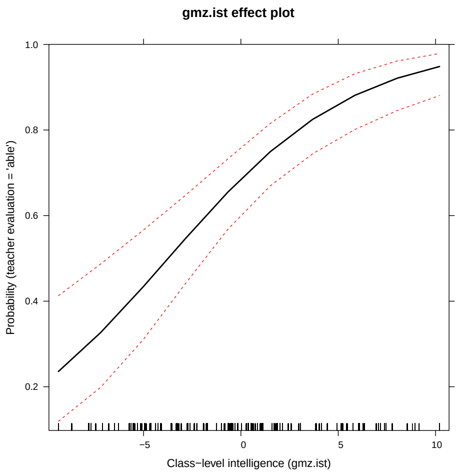


(e) Model 1e.

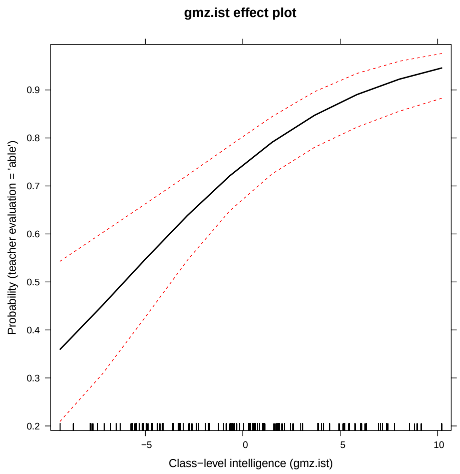


(f) Model 1f.

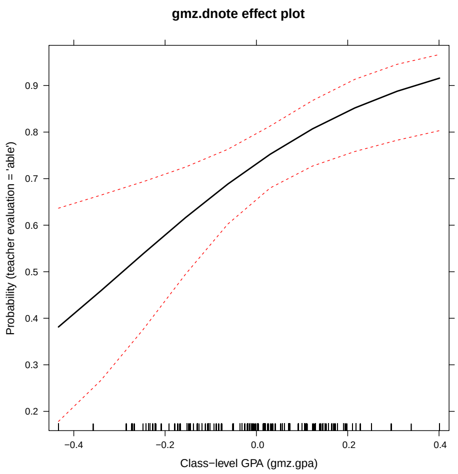
Figure A: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level achievement predictors in models 1a - 1f.



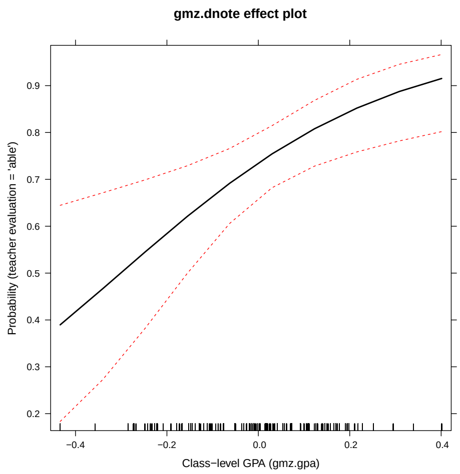
(a) Model 2a.



(b) Model 2b.

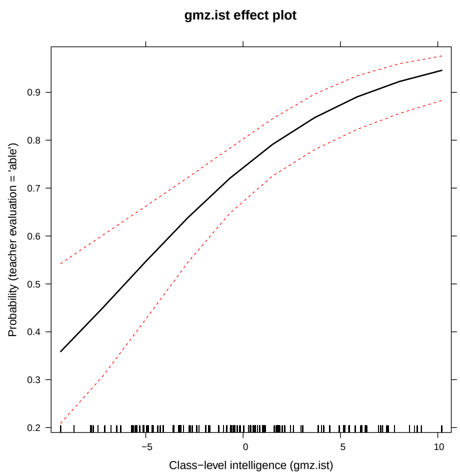


(c) Model 2c.

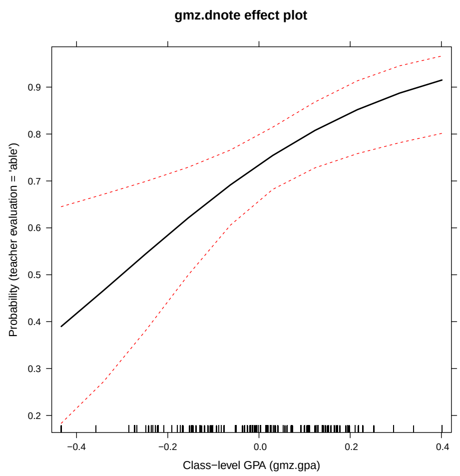


(d) Model 2d.

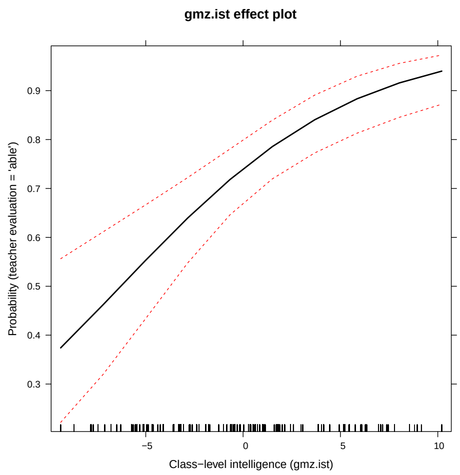
Figure B: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level achievement predictors in models 2a - 2d.



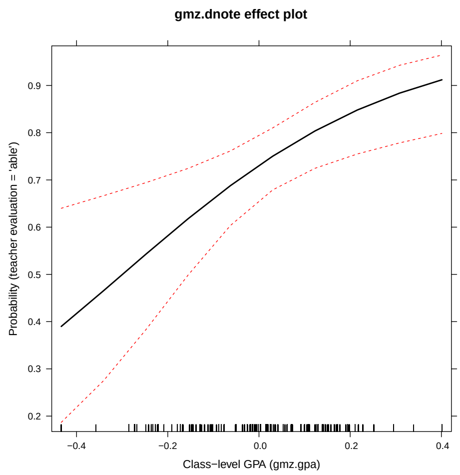
(a) Model 3a.



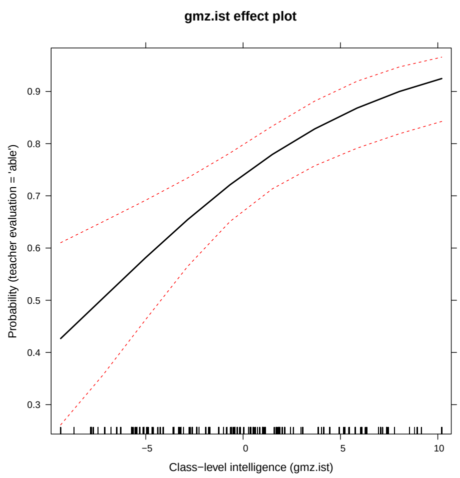
(d) Model 3d.



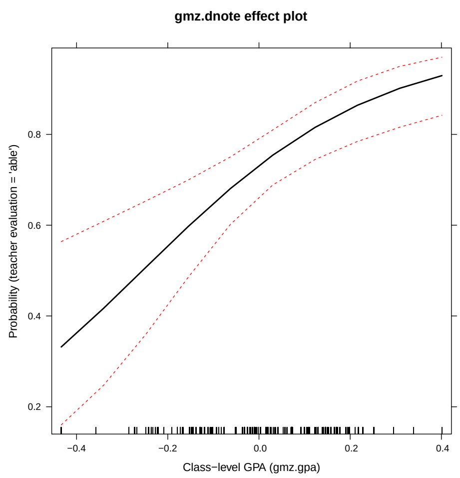
(b) Model 3b.



(e) Model 3e.

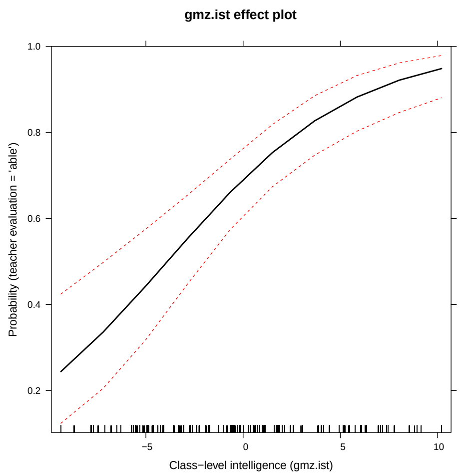


(c) Model 3c.

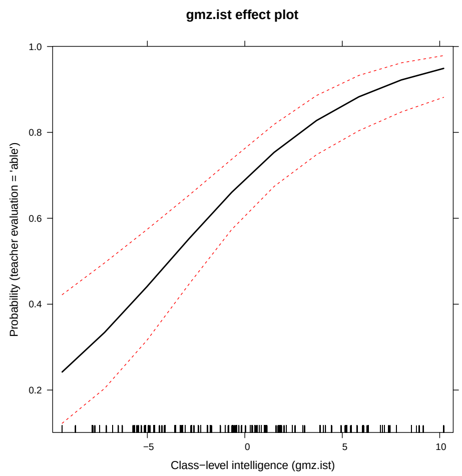


(f) Model 3f.

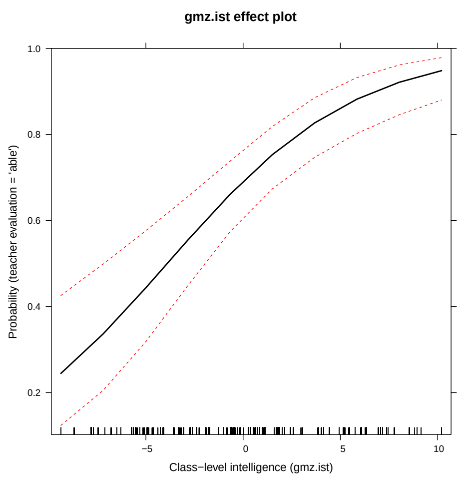
Figure C: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level achievement predictors in models 3a - 3f.



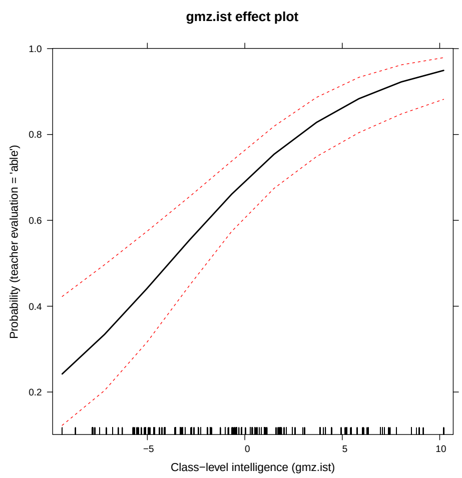
(a) Model 4a.



(b) Model 4b.

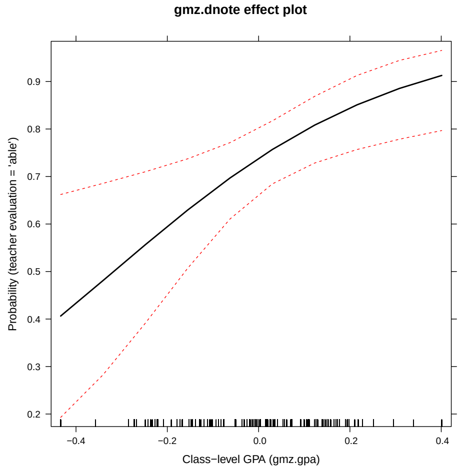


(c) Model 4c.

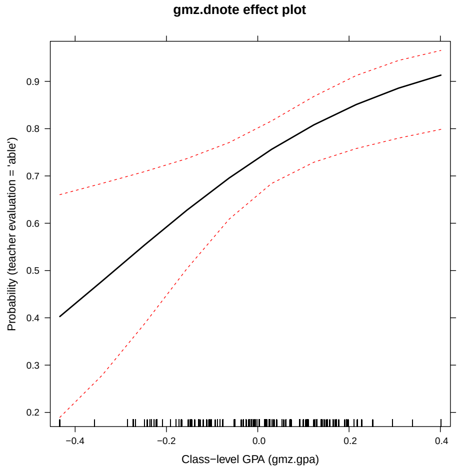


(d) Model 4d.

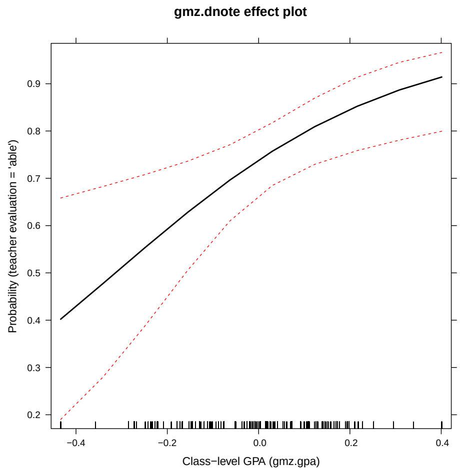
Figure D: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level intelligence in models 4a - 4d.



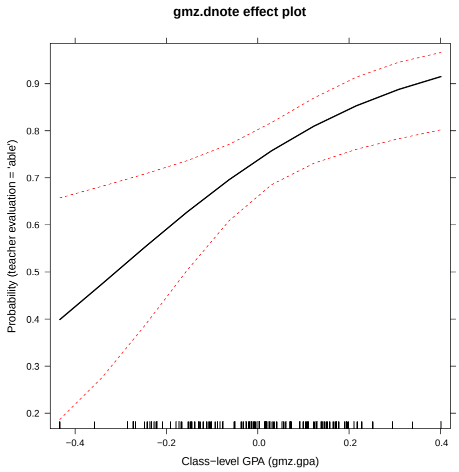
(a) Model 4e.



(b) Model 4f.

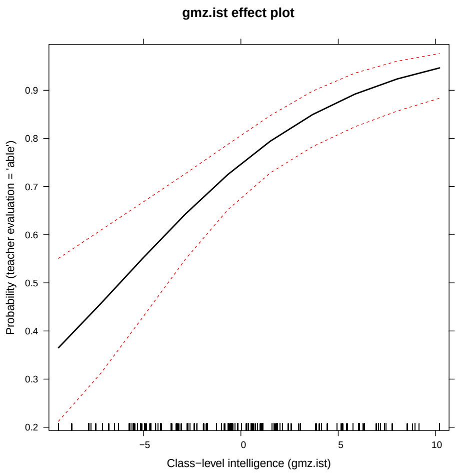


(c) Model 4g.

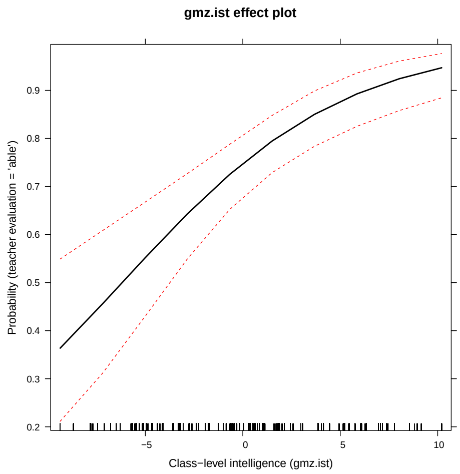


(d) Model 4h.

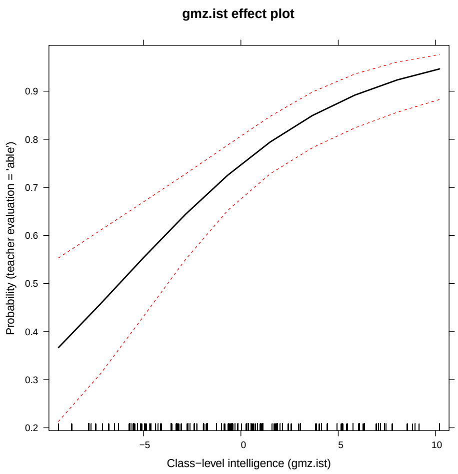
Figure E: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level GPA in models 4e - 4h.



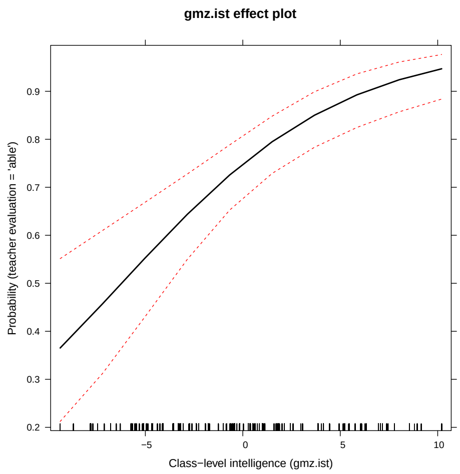
(a) Model 5a.



(b) Model 5b.

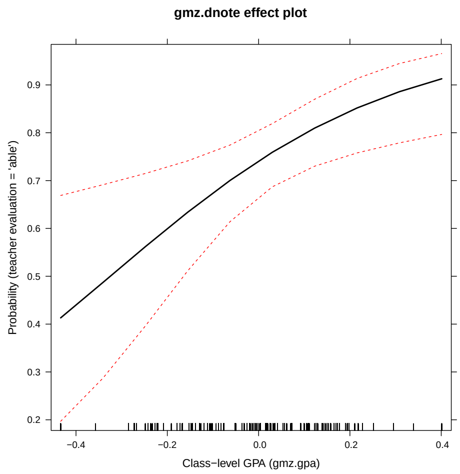


(c) Model 5c.

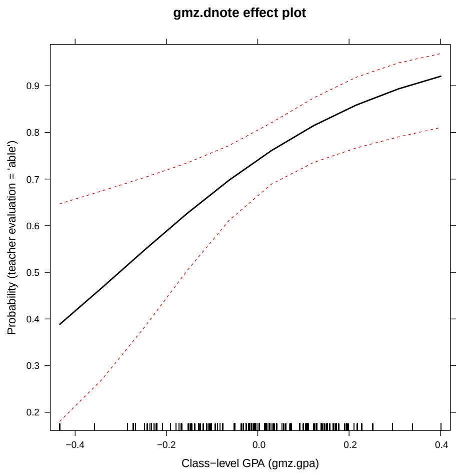


(d) Model 5d.

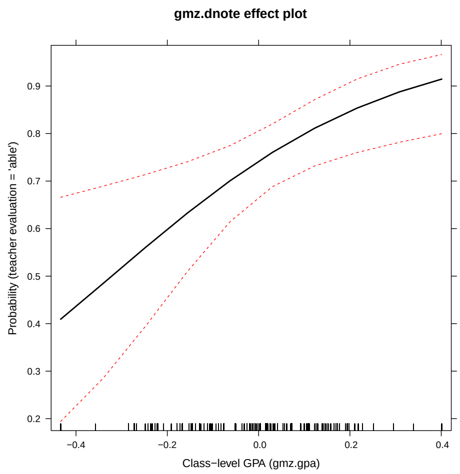
Figure F: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level intelligence in models 5a - 5d.



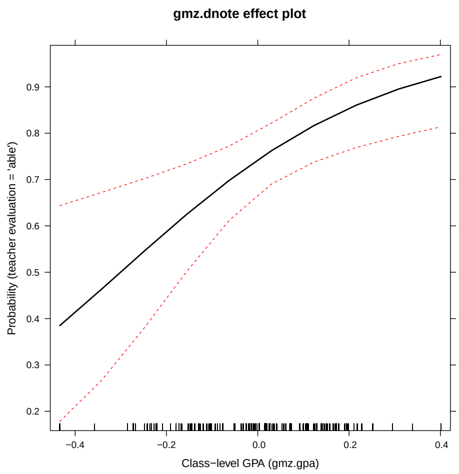
(a) Model 5e.



(b) Model 5f.

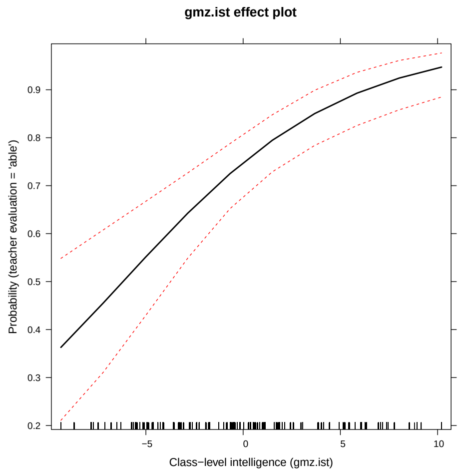


(c) Model 5g.

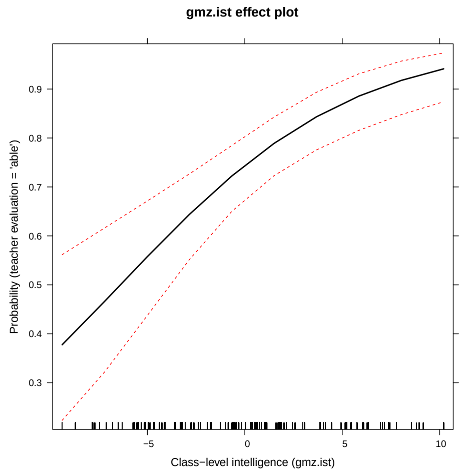


(d) Model 5h.

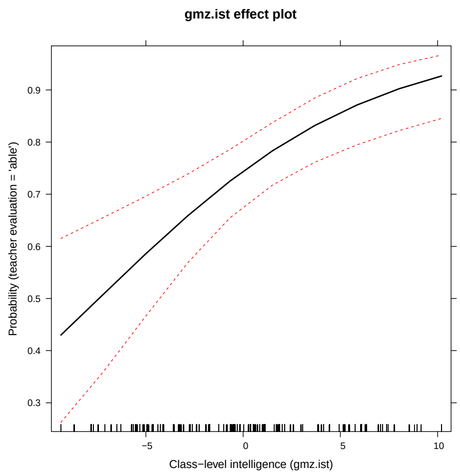
Figure G: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level GPA in models 5e - 5h.



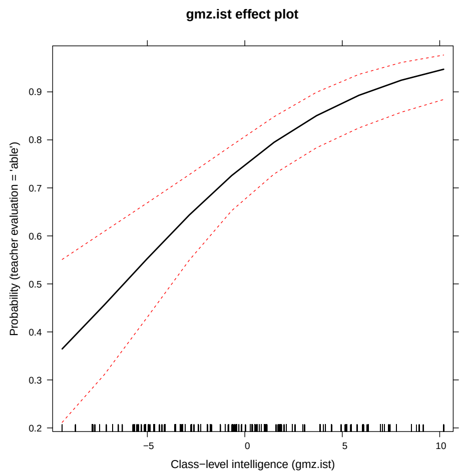
(a) Model 6a.



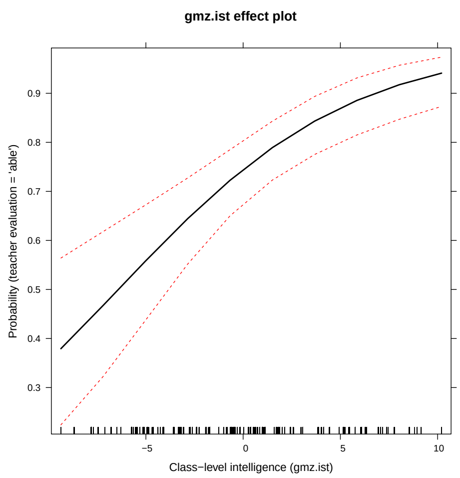
(b) Model 6b.



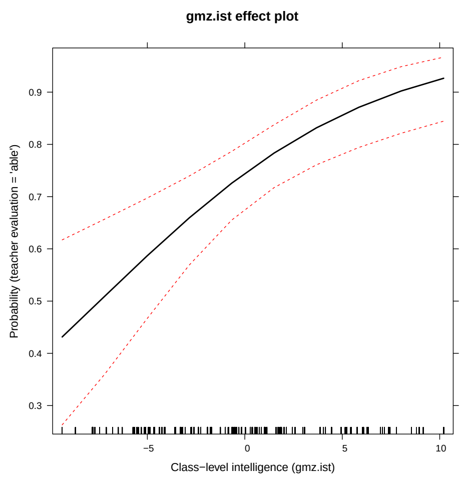
(c) Model 6c.



(d) Model 6d.

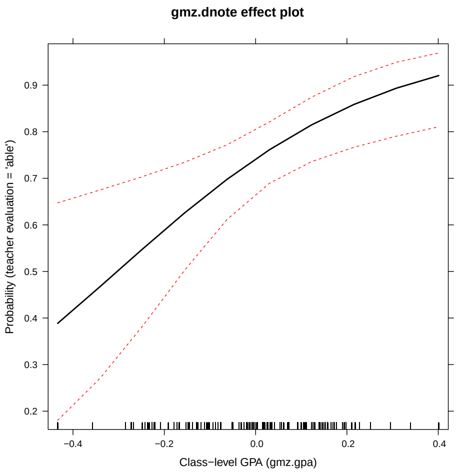


(e) Model 6e.

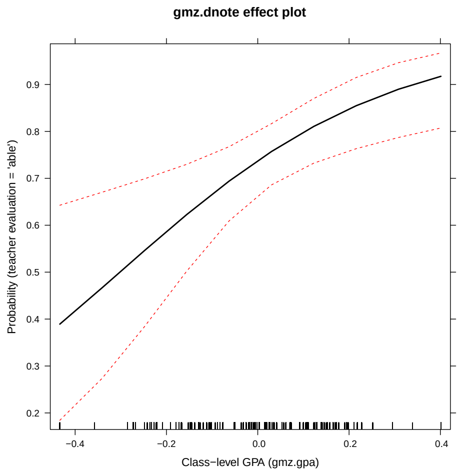


(f) Model 6f.

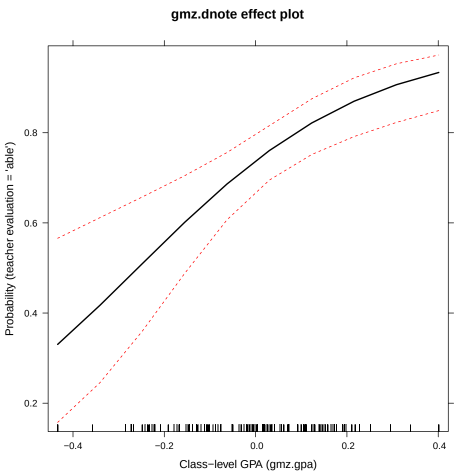
Figure H: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level intelligence in models 6a - gf.



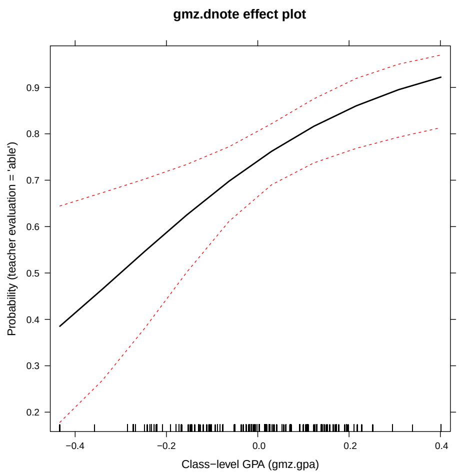
(a) Model 6g.



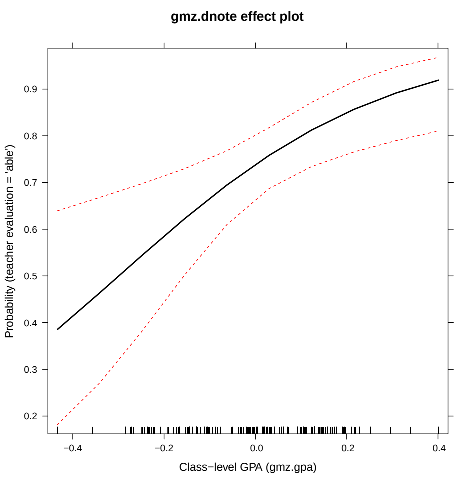
(b) Model 6h.



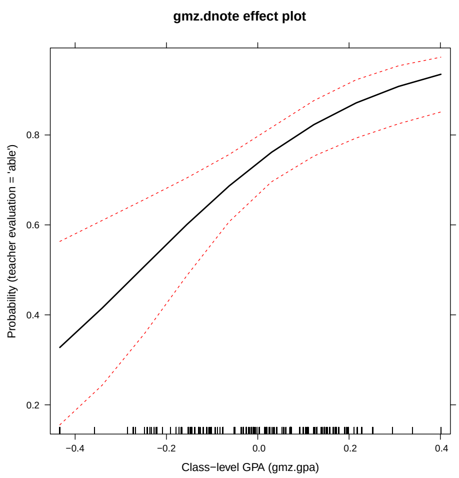
(c) Model 6i.



(d) Model 6j.

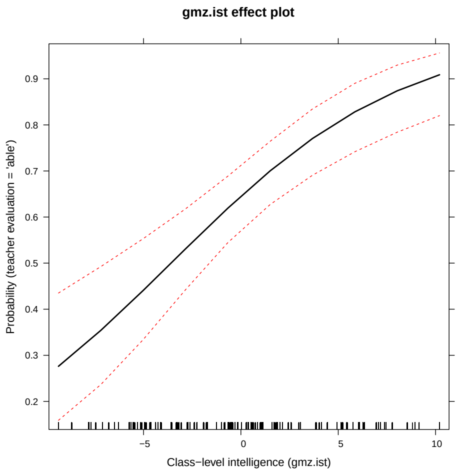


(e) Model 6k.

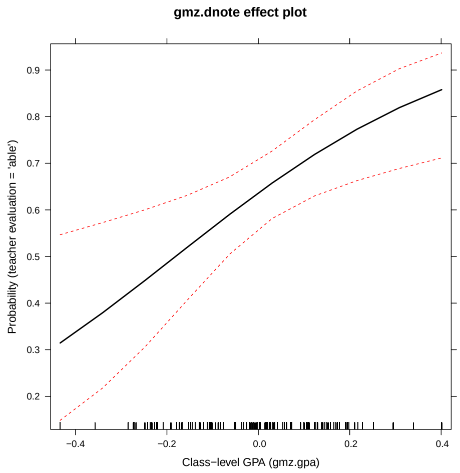


(f) Model 6l.

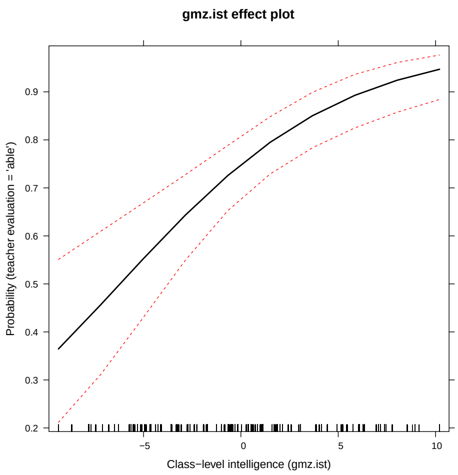
Figure I: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level GPA in models 6g - gl.



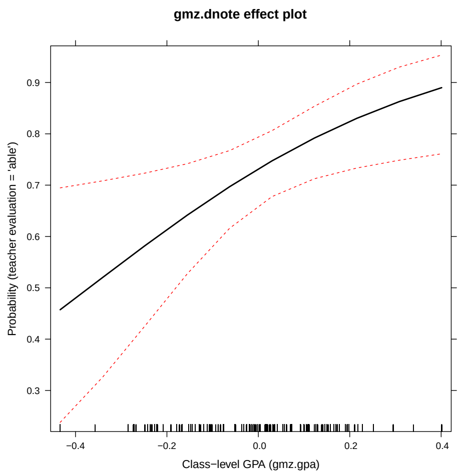
(a) Model 7b.



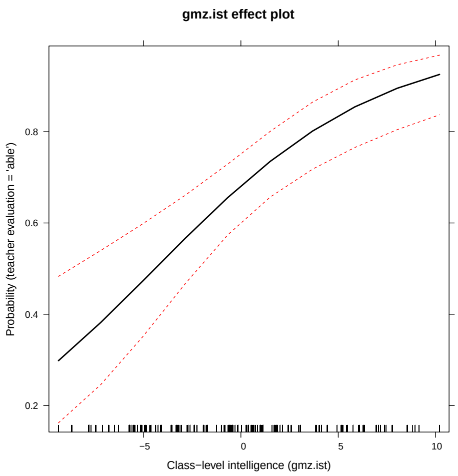
(b) Model 7c.



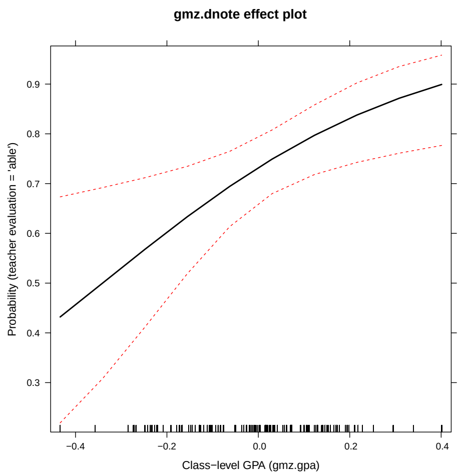
(c) Model 7d.



(d) Model 7e.

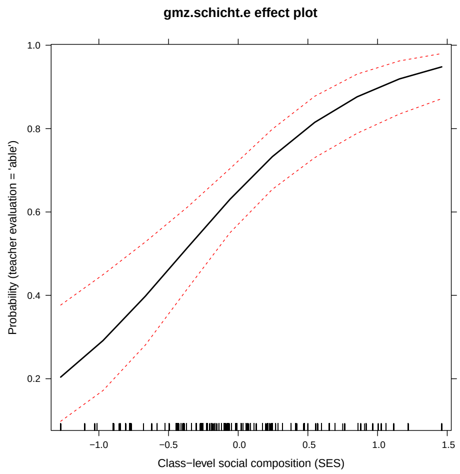


(e) Model 7f.

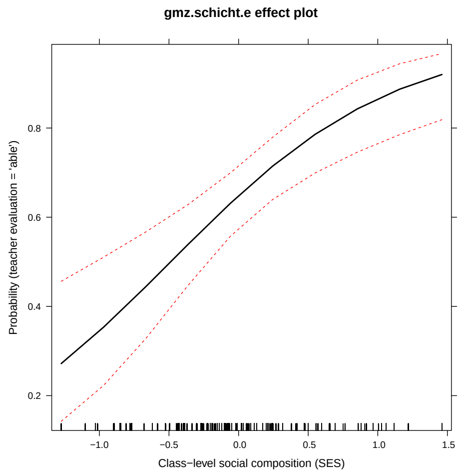


(f) Model 7g.

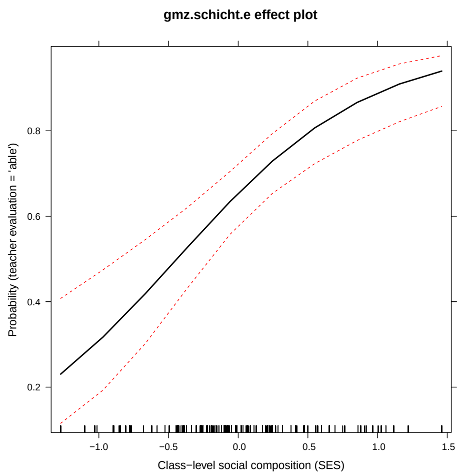
Figure J: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level achievement predictors in models 7b - 7g.



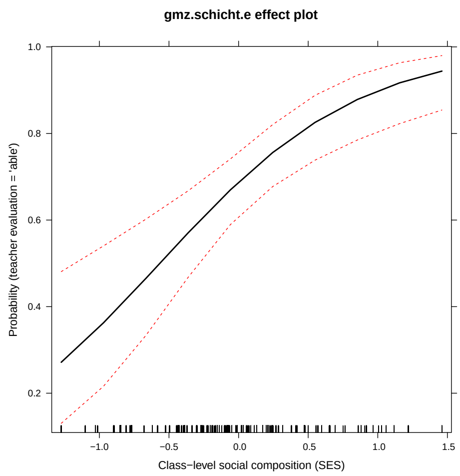
(a) Model 7a.



(b) Model 7b.

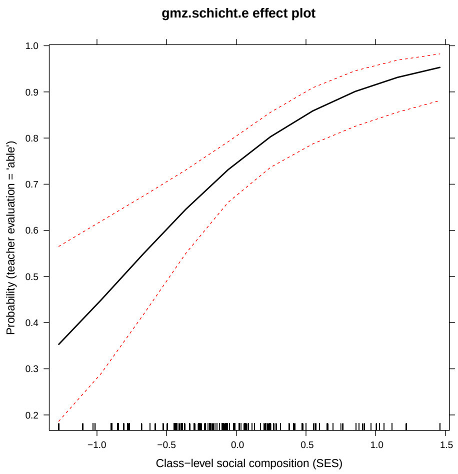


(c) Model 7c.

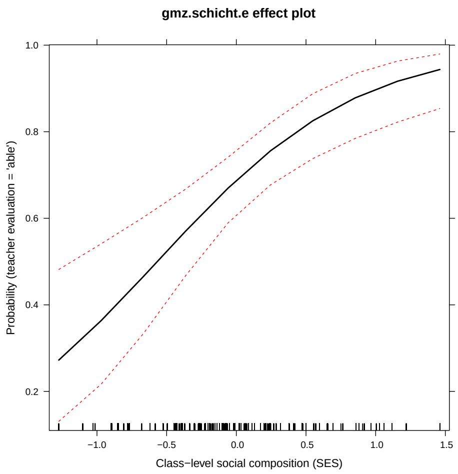


(d) Model 7d.

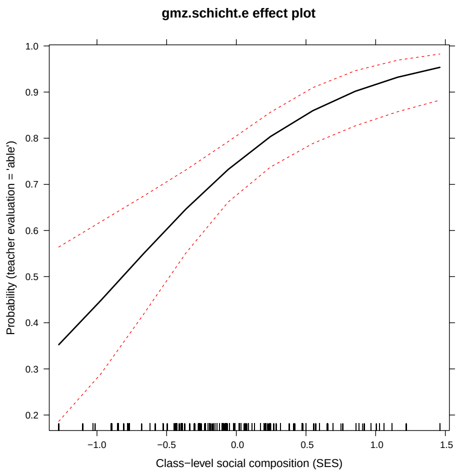
Figure K: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level social composition in models 7a - 7d.



(a) Model 7e.

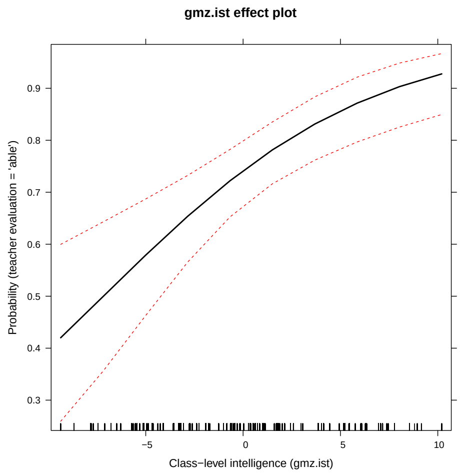


(b) Model 7f.

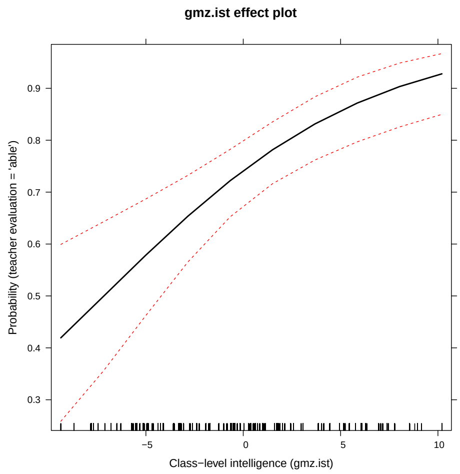


(c) Model 7g.

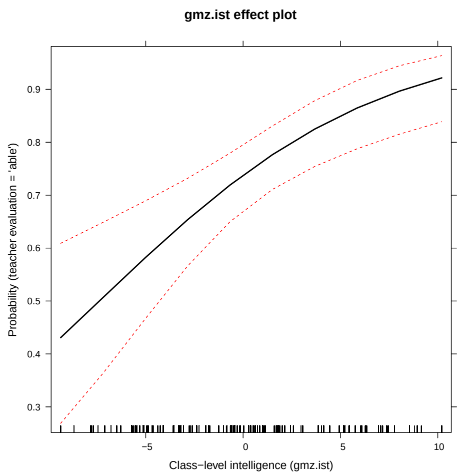
Figure L: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level social composition in models 7e - 7g.



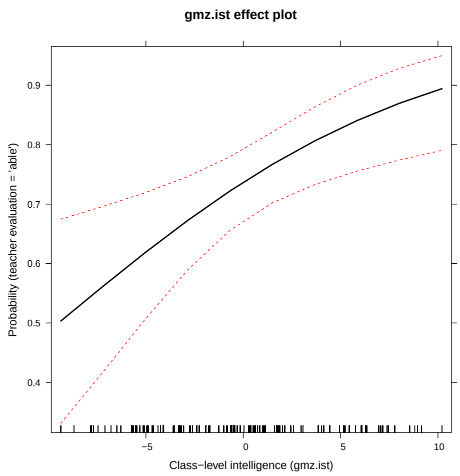
(a) Model 8a.



(b) Model 8b.

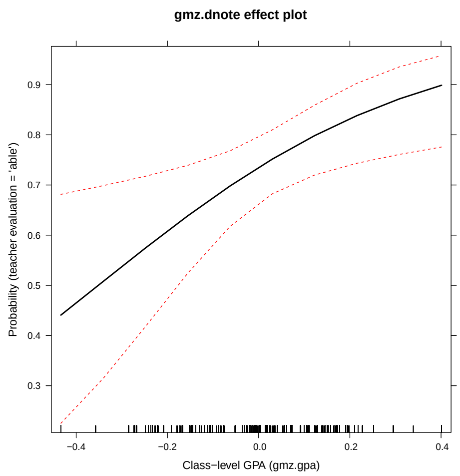


(c) Model 8c.

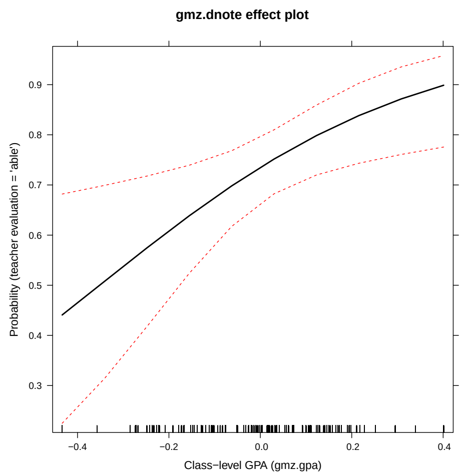


(d) Model 8d.

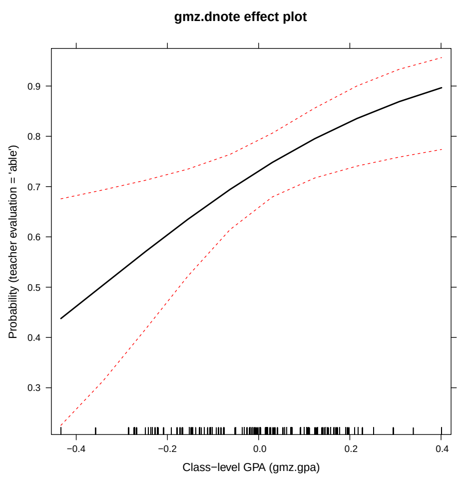
Figure M: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level intelligence in models 8a - 8d.



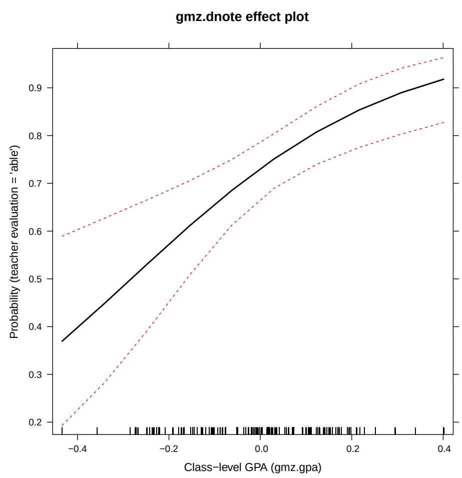
(a) Model 8e.



(b) Model 8f.

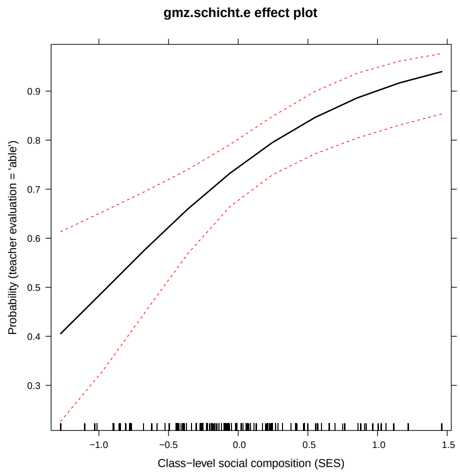


(c) Model 8g.

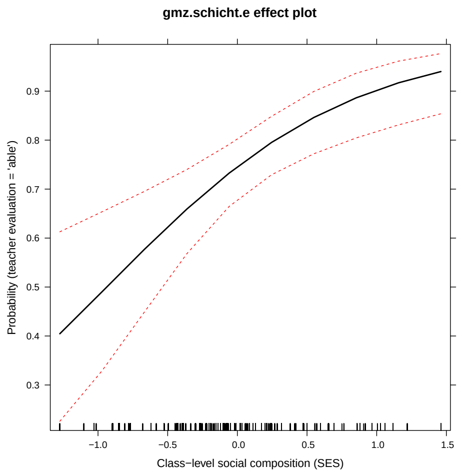


(d) Model 8h.

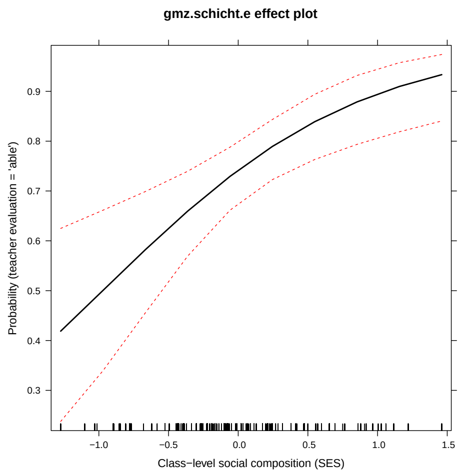
Figure N: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level GPA in models 8e - 8h.



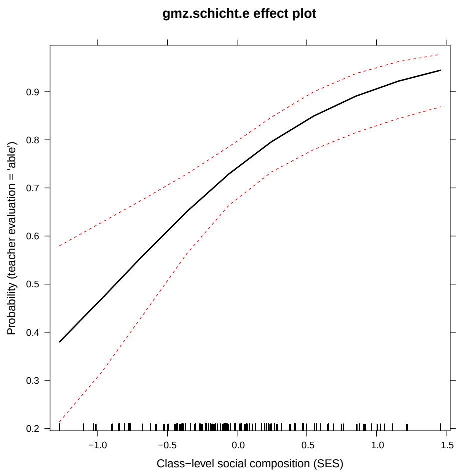
(a) Model 8a.



(b) Model 8b.

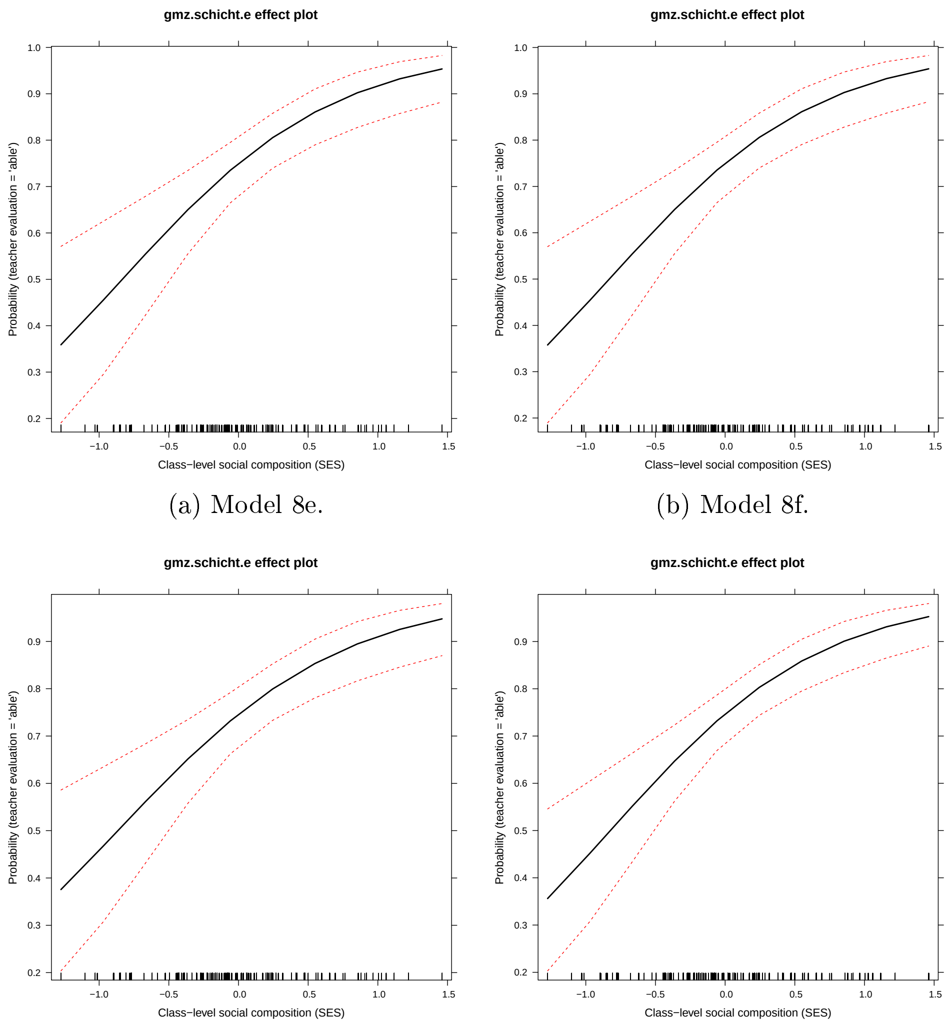


(c) Model 8c.



(d) Model 8d.

Figure O: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level social composition in models 8a - 8d.



(a) Model 8e. (b) Model 8f.

(c) Model 8g. (d) Model 8h.

Figure P: Average marginal effects (black solid line) and confidence intervals (red dashed lines) for class-level social composition in models 8e - 8h.

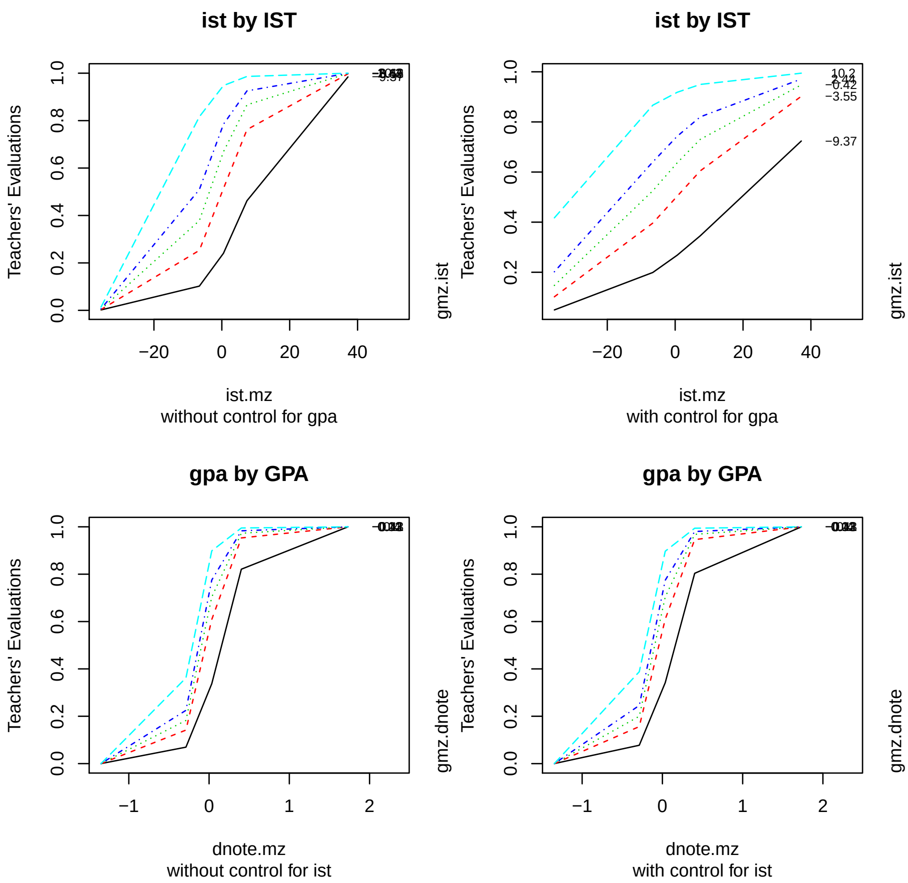


Figure Q: Marginal effects of cross-level ability interaction terms (alternative specification).

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

An Empirical Test of A Subjective-Expected-Utility Explanation

The objective of this paper is to integrate the idea of *Pygmalion* or self-fulfilling prophecy research (Rosenthal and Jacobson, 1968; Jussim and Harber, 2005) into the subjective expected utility framework of inequality in educational opportunities (Esser, 1999). In the theoretical section, a formal model about the impact of teachers' expectations on students' educational transitions in sense of a self-fulfilling prophecy is developed. In the empirical section, I test this model to predict both students' educational success (in terms of high school graduation) and their university transitions. Analyses control for both sample selection bias and unobserved heterogeneity. I find that in the underlying operationalization, teachers' expectations show significant effects on both educational success and university transitions. While the conditional decision problem of university transitions might lead to a selection bias, unobserved heterogeneity would have to be disturbingly high to affect the stability of self-fulfilling prophecy estimates.

1 Introduction

School surely is the first and by that way also the most important branching point in everybody's life course at least in industrialized countries. According to structural functionalists, the function of school is "to internalize in its pupils both the commitments and the capacities for successful performance of their future adult roles, and second (...) to allocate these human resources within the role-structure of the adult society" (Parsons, 1959, p. 298; also see Davis and Moore, 1945). 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, there is still room to refine social science theories on social *i*nequality in *e*ducational *o*pportunities (IEO).

On the one hand, powerful conclusions can be drawn from the theoretical framework that has been provided by social inequality theory based on rational-choice or *s*ubjective *e*xpected *u*tility (SEU) assumptions. One of its main strengths lies in allowing us to distinguish between primary and secondary effects of social inequality, i.e. between effects of socialization and effects of aspirations. Furthermore, SEU theory conventionally implies that research assumptions have to be formalized. This 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 underestimations and the *Golem* effect of self-fulfilling overestimations (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* or self-fulfilling prophecy research into the general subjective expected utility framework of IEO. In particular, I will refer to Esser's (1999) SEU-IEO that has been proposed in course of a concatenation of related theoretical accounts (Erikson and Jonsson, 1996; Goldthorpe, 1996; 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 I will focus on students' probability of achieving a high school degree, and on their propensity of beginning academic studies, respectively. The research design will follow the model proposed by Becker (2003) which includes controls for selection bias (Heckman, 1979). Additionally, I will perform a sensitivity analysis for all self-fulfilling prophecy indicators to test their robustness against a vector of unobserved covariates (Buis, 2007, 2010, 2011).

This paper will be structured as follows: First, the basic assumptions of both the SEU-IEO model and *Pygmalion* will be discussed. Then I will outline to what extent *Pygmalion*'s implications require us 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 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 in Educational Opportunities: Educational Transition Models

One of the most influential theoretical components of Boudon's (1974, p. 29ff.) monograph for contemporary quantitatively-oriented IEO research is its distinction between primary and secondary effects of social inequality.

The *primary* effect of educational inequality states that the lower educational success of lower-SES children may be due to their lower capabilities – be they defined as educational interests, intellectual skills, effort, or motivation (Jackson et al., 2007; also see Müller-Benedict, 2007). *Part* of the primary effect may in fact be genetic, but another, presumably greater part of the above-mentioned characteristics is acquired during socialization (Erikson and Jonsson, 1996, p. 10f.).

The secondary effect, contrarily, operates via stratum-specific differences in educational decision making due to differential opportunity-cost structures, and Boudon's crucial assumption is that secondary effects still take place once primary effects have been controlled for (Boudon, 1974, p. 29ff.; for a critique see Nash, 2003).[1] The idea that utility considerations may shape students' (or their parents') educational decisions was taken on in a series of consecutive theoretical models proposed by Goldthorpe (1996), Erikson and Jonsson (1996), Breen and Goldthorpe (1997) as well as Esser (1999). A common proposition of these models relates to the idea that it is simpler (i.e. more parsimonious) to assume that there is no class-specific variation in either aspirations towards education *per se* or in potentially underlying value systems. Instead, education is regarded as an investment good the costs and benefits of which vary by social classes. Each family will strive to avoid downward mobility, but unsurprisingly, for lower-educated parents, this goal will be reached already for lower educational qualifications of their children – while for higher-educated parents, a far higher degree will have to be obtained. Moreover, for the offspring of parents in less advantageous positions, each failed attempt of trying a higher educational alternative will be more serious in its consequences concerning both monetary (earnings foregone; loss of financial support) and transactional costs (a loss in itself; the risk of dropping out of the educational system).

Erikson and Jonsson (1996) introduce a simple 3-parameter model postulating that students' utility of continuing education (or opting for the comparably higher educational track) can be regarded as a function of the product of educational benefit B and the expected probability of educational success P, minus expected costs of education C ($U = PB - C$). Esser (1999, p. 165-175) takes up these crucial parameters in order to develop a model according to the logic of subjective expected utility theory. This more complex model is elaborated on in the next paragraph.[2]

[1] Some authors also subsume class-sensitive structural conditions at the school level – e.g. in terms of the variance of teachers' school track recommendations by parental social class (Pietsch and Stubbe, 2007) among secondary effects of social inequality as well (e.g. Müller-Benedict, 2007).

[2] While the model by Erikson and Jonsson (1996) shows the highest theoretical proximity to the Esser (1999) model, also the theoretical accounts by Meulemann (1979) and Breen and Goldthorpe (1997)

Esser's' Subjective-Expected-Utility model Esser (1999) uses a *s*ubjective-*e*xpected-*u*tility (henceforth referred to as *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:

$$EU(A_n) = P_{sd}(-SD) \tag{1}$$

$$EU(A_b) = P_{ep}B + (1 - P_{ep})P_{sd}(-SD) - C. \tag{2}$$

Here, SD is the value of status decline with P_{sd} as its impact (in terms of a subjective probability) on parental decisions; B is the benefit of higher education (e.g. in terms of labour market prospects); P_{ep} is the subjective probability of successfully completing the chosen school track; and C are the expected costs of education (also see Becker, 2003; Pietsch and Stubbe, 2007). By simple linear transformations, Esser (1999) shows that

$$EU(A_b) > EU(A_n) \iff B + P_{sd}SD > C/P_{ep}, \tag{3}$$

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. Since in case of low P_{ep}, educational motivation has to be very high to exceed the critical threshold of the investment risk, the model can also account for the fact of persisting inequality in educational opportunities (Esser, 1999, p. 270).

Among educational transition models, both the Breen-Goldthorpe- and the Esser model have been tested most comprehensively (Jonsson, 1999; Breen and Jonsson, 2000; Becker, 2003; Stocké, 2007; Schneider, 2008). However, in terms of methodology, Becker's (2003) operationalization controlling for selection bias via "Heckit" correction (Heckman, 1979) was referred to being the best available test of Esser's (1999) model (Stocké, 2007, p. 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 the 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 for the presence of another endogeneity that apparently has been neglected so far but is worth considering: 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 first established by Robert Merton (1948). In this seminal 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)

evidently capture similar ideas.

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 students future school achievement has been labeled as the *Pygmalion* effect.[3] The idea behind the metaphor holds that teachers' too high or too low expectations can have an impact on the teacher-student interaction, which, in turn, might influence the students to adopt their motivations and aspirations according to their teachers' expectations. In the words of Merton, teachers' expectations, which had originally been misled, would in turn lead 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 claimed 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. Although the 'late bloomers' were actually selected randomly, Rosenthal and Jacobson (1968) observed that in an IQ test which was administered one year later they gained significantly better test scores than the control-group students. Thus, the false expectations of the teachers (who had been led to believe in the artificially created group of late bloomers) had become true.[4] 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).[5]

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, psycho-physical 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).[6] A second meta-analysis based on a more narrowly defined set

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

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

[5] 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).

[6] Both d and r are measures of meta-analytic effect sizes. Effect size r can be obtained from t, F, χ^2, and Z statistics. Effect size d is a linear transformation of r. For a more elaborate discussion see

of *Pygmalion* studies examined 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 additionally revealed that the effect of teachers' expectations at t_0 on later IQ scores at t_1 highly depends on how long the teachers have already been teaching a particular class.[7]

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 mediators and 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 either mediators or moderators.[8] In particular, more research is obviously needed with regard to students' social backgrounds (Jussim and Harber, 2005). Concretely, there are only three studies who explicitly considered these effects: First, Madon et al. (1997) found that self-fulfilling prophecies appear to be stronger among students who had a 'prior history of low-achievement', which was operationalized as their standardized results in a test that had been administered prior to the actual experiment. Basically, the authors' operationalization of self-fulfilling prophecies as teachers' over- and underestimations appears to be very promising (and will thus also be used in the study at hand): Whereas in experimental studies such as the original *Pygmalion* study (Rosenthal and Jacobson, 1968), the researcher can expose a teacher to false information in order to ensure that her expectations are really inaccurate, in naturalistic (such as survey data) studies, this is not equally possible (Jussim, 1986). To overcome this problem, Madon et al. (1997) first regressed teachers' expectations (related to students' performance, talent and effort) on a set of student background variables. In a second step, they used the residuals of these regressions – reflecting a student's over- or underestimation by her teacher – as new variables to ensure that a teacher's expectation is, to some extent, actually inaccurate. However, for their purpose of identifying moderator variables their model unfortunately suffers from a methodological weakness.[9]

Second, Jussim et al. (1996) found evidence that self-fulfilling prophecies are moder-

Rosenthal and DiMatteo (2001).

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

[8] Among the few exceptions of empirical studies taking moderator effects into account, the meta-analysis by Raudenbush (1984) which found that the effect size of self-fulfilling prophecies varies by 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).

[9] Concretely, among the set of background variables that was used to identify teachers' over- and underestimations, we can find students' 5^{th} grade math test scores – which were also used to identify low and high achievers (Madon et al., 1997, p. 798). Therefore, it is not surprising that the authors find a variation in the effect size of self-fulfilling prophecies based on this variable.

ated by both social class and ethnicity-related variables. In their study, the standardized relationship between teachers' perceptions and students' future test scores was about .25 for students of parents with a lower education, and .03 for students of parents with a higher education. In the United States, similar differences could be detected between Caucasian students and African-American ones in terms of a standardized effect size of .14 and .37, respectively. Third, Madon et al. (1998) noted that teachers' perceptions of students' performance and talent (but not of 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 such as students' school grades, intelligence test scores and their motivation are introduced in multivariate analyses. Hence, the largest share of the differences that teachers identify between social groups corresponds closely to actual differences in previous 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 (Rosenthal and Jacobson, 1968). Second, one 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. This is precisely where the SEU-IEO framework comes back in: Just as self-fulfilling prophecy research depends on considering student background variables, the SEU-IEO framework lacks the consideration of exactly that endogeneity concerning students' probability of educational success which is the main point of all *Pygmalion* studies. The task in 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 outlined in section 2.1, educational decisions would be a direct function of net utility. However, this seems to be only half 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 students' 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 also assume an endogeneity of p_{ep}[10], i.e. of the subjective expected probability of successfully completing the chosen school track. Following both Breen and Goldthorpe (1997, p. 285) as well as Esser (1999, p. 272f.), the subjective probability of educational success depends on students' objective school performance. In accordance with Esser's notation, this reads

$$p_{ep} = f(AP), \tag{4}$$

while AP denotes students' academic performance. Claiming that teachers' expectations in terms of a 'net' effect of self-fulfilling prophecies (Madon et al., 1997) at time t, TE_t, may influence students' academic outcomes at a later time $t+1$ can be formalized as

$$AP_{t+1} = g(TE_t). \tag{5}$$

For $p_{ep_{t+1}}$ thus holds

$$p_{ep_{t+1}} = f(g(TE_t)) \tag{6}$$

– meaning that subjective probability assumptions are a function of students' objective school performance which is, in terms of a self-fulfilling prophecy, dependent on teachers' earlier expectations. Notably, in self-fulfilling prophecy research, many studies stress that the crucial mechanism of teacher expectancy effects also operates via students' self-concept and their aspirations (Jussim, 1989; Gill and Reynolds, 1999; Muller et al., 1999; Mechtenberg, 2009; Mistry et al., 2009).[11]

Will AP_{t+1} be the only variable that affects $p_{ep_{t+1}}$? Certainly not. Concretely, I assume that equation (6) can be decomposed into

$$p_{ep_{t+1}} = h(p_{ep_t}, AP_{t+1}, \epsilon). \tag{7}$$

Equation (7) expresses that students' subjective expected probability of educational success is a function of her preceding subjective expected success probability, her actual academic performance and an unspecified teacher treatment effect ϵ that captures classroom praise, bilateral encouragement, and similar mechanisms (without making any assumptions about the functional form of this relation).

We should now apply this idea on Esser's (1999) formal model by tracking the logic of a SFP in its appropriate survey-data framework of teachers' over- and underestimations.[12] Let $\delta \in \{0, 1\}$ indicate whether a student has been underestimated ($\delta = 0$) or overestimated ($\delta = 1$) by her teacher. Restricting the other SEU parameters to remain constant over time, let further $\tilde{p}_{ep_\bullet} = p_{ep_t} + \Delta p_{ep}$ to get rid of time indices (see Jaeger and Holm,

[10] To appreciate that this term and p_{sd} refer to subjective expected probabilities, I will use lowercase p in the following.

[11] **?** propose a formal model of *Pygmalion* effects applied to management science which also highlights the importance of subjective success probability assumptions; but since this model is neither tangent to the field of educational transitions nor to the respective rational action model, I do not discuss it in more detail.

[12] In section 3.3, I will measure teachers' over- and underestimations by the residuals of a regression of teachers' evaluations on both students' performance and their motivation, (see Madon et al., 1997).

2011, for a similar analytic strategy) – that is, $\tilde{p}_{ep_\bullet}$ captures students' initial subjective expected probability of educational success *plus* (or minus) the additional gain (or loss) Δp_{ep} that is due to an over- (or under-)estimation as sketched in (7).[13] If finally $\tilde{p}_{ep_-}$ and $\tilde{p}_{ep_+}$ denote the under- and overestimated students' corresponding subjective expected probability of successfully completing the chosen school track, respectively, then we can rewrite equation (3) as follows:

$$EU(A_b) > EU(A_n) \iff B + p_{sd}SD > \frac{C}{[\delta \cdot \tilde{p}_{ep_+} + (1-\delta) \cdot \tilde{p}_{ep_-}]} \tag{8}$$

As the argument goes, on average $p_{ep_{t+1|\delta=1}} > p_{ep_{t+1|\delta=0}}$ since on the one hand, $AP_{t+1|\delta=1} > AP_{t+1|\delta=0}$, and on the other hand, $\epsilon_{\delta=1} > \epsilon_{\delta=0}$. Holding constant for p_{ep_t}, it follows that $\tilde{p}_{ep_+} > \tilde{p}_{ep_-}$. Since the denominator of the right-hand side of equation (8) then is larger for $\delta = 1$ than for $\delta = 0$, it follows that

$$EU_{\delta=1}(A_b) > EU_{\delta=0}(A_b), \tag{9}$$

that is, other things being equal, students who had been overestimated by their teachers should have a higher expected utility of choosing the next higher-level school-track than students who had been underestimated.

Model identification In the above-sketched model I assume that self-fulfilling prophecies directly enter the students' utility function. However, it has to be clarified which type of rationality students' utility function relies on. A student who has been overestimated by her teacher will dispose of a higher subjective expected probability of educational success not only because of her better grades, but also because of more subtle teacher treatment effects (above referred to as ϵ) that may be understood quite similarly to the 'caress' effect in Ovid's metamorphoses. As Morgan (1998, p. 136) writes, students "adopt the expectations that others have of them and add these to their own expectations formed independently through their own rational self-reflection". Adding expectations of teachers in their role as *significant others* (Sewell et al., 1969, 1970; Morgan, 1998, 2002) may be an unconscious endeavor, but in altering a crucial parameter of the utility function, they might also affect quite rational utility considerations.

The question is now how the crucial parameters can be estimated in an empirical model. My answer is that the current framework in social psychology of *residualizing* teachers' expectations (Madon et al., 1997, 2006) is very helpful for identification purposes. The assumption is that when teachers' expectations are residualized for students' achievement, their motivation and self-concept at t_1, the relations that are addressed in (7) can be approximated also in case of lacking empirical measures. While p_{ep_t} can be measured from the data at hand (see section 3), I do not have indicators for AP_{t+1} and ϵ. But I assume that by regressing teachers' evaluations on a set of performance and motivation variables, *differences* in students' academic performance between t and $t+1$

[13] To some extent, my theoretical model resembles the *value-added* approach in school effectiveness research (Ladd and Walsh, 2002; Rivkin et al., 2005).

as well as unobserved teacher treatment factors that exceed (or undershoot) students' motivation and their self-concept at t (and thereby take effect as a causative factor for Δp_{ep}) can, if admittedly not entirely isolated, at least be approximated.

Hence, I assume that residualizing teachers' expectations as proposed by Madon et al. (1997, 2006) provides a useful tool for approximating the unobserved mechanisms that enter students' rational utility function by both consciously and unconsciously influencing a crucial parameter thereof.

2.4 Hypotheses

After these theoretical considerations my main hypothesis is easily outlined: I postulate that via their prognostic nature, teachers' expectations have distinct effects on students' educational transitions in terms of a self-fulfilling prophecy. By 'distinct effects' I mean that they will have a significant impact apart from the convenient theoretical concepts of the SEU-IEO model (Esser, 1999). Since a major claim of rational-action theories of educational transitions 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 (Breen and Goldthorpe, 1997), as a first step I aim to analyze the probability of German 10^{th} class *Gymnasium* students to achieve a high school degree ('Abitur'). This certificate all along used to be and still is the crucial prerequisite for access to tertiary education. When surveyed mid of 10^{th} grade, German *Gymnasium* students are still facing a crucial decision: They could continue education in secondary school level II ('Gymnasiale Oberstufe') that would be successfully finished by obtaining Abitur after (at that time consistently) three years – or they could quit secondary school immediately at the end of 10^{th} grade, or thereafter, without having passed Abitur. As Schneider (2008, figure 2b) recently observed by using *GSOEP* data, even after 9 years of secondary-school education, for students from the salariat, the survivor function modeling the probability of not dropping out from *Gymnasium* lies remarkably above the corresponding survivor function for students from working class. Hence, one can assume that secondary school cost-benefit considerations as postulated by Esser (1999), *inter alia*, are equally an issue for passing Abitur.

In a second step, I will also model students' transition propensities to tertiary education in terms of starting academic studies. Becker and Hecken (2009a,b) argue that utility considerations as reflected in Esser's (1999) SEU model are also pivotal for university transitions. One social mechanism that could account for transition differences between the social strata at this comparably later point in students' educational life course is that their respective *time horizon* might differ, too. The impending costs of higher education are accompanied by merely uncertain returns – which may be more significant for students from the lower social classes than for those from the salariat (Becker and Hecken, 2009b, p. 235f.). Therefore, the former can be expected to make that transition less frequently than the latter.

Due to the fact that rational-action theories on IEO in general have proven useful also in predicting higher-level transitions (Jonsson, 1999; Need and De Jong, 2001; Becker and Hecken, 2009a,b; Hillmert and Jacob, 2010), I take the SEU model as given in order to keep the number of hypotheses reasonably small for avoiding difficulties in causal identification of my models. Therefore, according to the formal model of self-fulfilling prophecies I have proposed above, there remain only two (nonetheless important) hypotheses to test:

H_1 : Apart from the SEU-model indicators, students' probability of achieving a high school degree increases with (positive) self-fulfilling prophecies, SFP.

H_2 : Apart from the SEU-model indicators, students' probability of starting academic studies increases with (positive) self-fulfilling prophecies, SFP.

As indicated above, due to data restrictions, I am not able to test for a direct impact of teacher expectations on students' future school performance (as *Pygmalion* in its initial form would require). However, I assume that given the (in my view) adequate operationalization of SFP in terms of over- and underestimations – approximating the factors that affect student differences in their subjective expected probability of educational success Δp_{ep}–, we can identify an estimate that gets quite close to the unobserved mechanisms of the 'real' self-fulfilling prophecy.[14] The next section will provide an insight into which measures will be used concretely.

3 Operationalization

3.1 Data

All analyses will be based on a German panel dataset which is known as the 'Kölner Gymnasiasten-Panel' (Engl. '*C*ologne *Hi*gh *S*chool *P*anel', henceforth referred to as *CHiSP*). The CHiSP consists of an initial (student-level) survey from 1969 (Gesis-No.: ZA0600) with $N = 3385$ 10^{th}-grade *Gymnasium*[15] students in North Rhine-Westphalia with three re-surveys in 1985 (Gesis-No.: ZA1441; $N = 1987$), 1996/97 (Gesis-No.: ZA4228; N=1596), and 2010 ($N = 1301$; no Gesis-No. available yet). In the initial survey, students were asked about issues like their performance, interests and plans in school and about their social background and their relationship to their parents. Simultaneously to the initial survey, the students took part in an *I*ntelligence *S*tructure *T*est (IST) containing four sub-scales 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

[14] 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 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).

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

about their social background, their style of raising children and their aspirations for their children. Amongst others, teachers were asked about several 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 provided detailed information on their educational and occupational careers until the age of 43. I chose this admittedly older data, because to the best of my knowledge, it is the only available longitudinal dataset that contains appropriate measures of both indicators of the SEU model outlined by Esser (1999) and of teachers' expectations that are required to construct over- and underestimations in order to operationalize self-fulfilling prophecies adequately. The latter indicator will be described in the next but one paragraph.

3.2 Variables

Dependent variables In the hypotheses section I identified two dependent variables. The first dependent variable is defined by whether 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, I 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, I want to focus on those students only who passed Abitur on the first try (event=1) using all observations that did not achieve Abitur within 3 years after the 10^{th} class survey in 1969 as a reference (event=0).[16] The second dependent variable is given by whether the former students have ever started academic studies. Since my analyses will be based on panel data, I have to take into account that from a theoretical point of view, it would be possible for the former students to start academic studies at any later point in time – including data points set after the last survey of the CHiSP (currently 2010). 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 were to be considered a necessity in order for them to reach their aim in life. Students had the following reply options: 1 'yes, necessary'; 2 'useful, but not necessary'; and 3 'not important'. I 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 their child did not pass Abitur. The categories of this variables are 1 'not much'; 2 'little'; 3 'very disappointed'; 4 'would be the worst'. I dichotomized this variable as follows: 0 'not much / little'; 1 'very disappointed / would be the worst'. I operationalize the *expected status decline*,

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

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 offspring is 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' (see Becker, 2003, for a comparable operationalization of the SEU predictors).[17] To keep results comparable with previous tests of the SEU model (Becker, 2003; Becker and Hecken, 2009a,b), I follow these authors in presenting both an additive SEU model as well as the interaction terms for students' educational motivation and their investment risk as required by the Esser (1999) model.

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 whom students they suppose to be able for academic studies, and whom of them not. Since the question was phrased openly, 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, available upon request). 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 were to loose substantive information when proceeding on this assumption.

To overcome the first problem, analyses reported below will focus on class teachers' evaluations only. I expect that the intra-individual variance of teachers' evaluations partially depends on the quality of teacher-student relationships. I assume that class teachers have a more intense relationship to and a better knowledge of their students than 'ordinary' teachers. Thus, looking only at class teachers' evaluations will both simplify the data structure and overcome the problem of inter-teacher variance.[18] In order to overcome the second problem, as a preliminary analysis, two logistic regressions

[17] Dichotomization of all SEU predictors follows both a theoretical and a methodological directive. Regarding theory, the SEU model explicitly demands certain terms such as p_{ep} or p_{sd} to be 0-1 coded (Esser, 1999, p. 269). Regarding methodology, the application of techniques correcting for sample selection bias (Heckman, 1979) requires to estimate a probit model in the first step – necessitating a dichotomous outcome.

[18] As regards social mechanisms, I 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 in a "grading continuation game" (Mechtenberg, 2009, p. 1437f.).

of the chances of getting a positive evaluation vs. getting a negative one, or none at all, on students' intelligence, average grade, social background, motivation and gender were estimated (not shown, available upon request). These results indicate that for the chances of getting a positive evaluation vs. not getting one at all, the effect sizes of all independent variables point to the same direction, but they are notably lower than for the chances 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 they score *higher* than students with a bad teacher evaluation. However, in the analyses presented below I will only 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 of them z-transformed). The motivational component comprises students' subjective assessments of i) their homework effort, ii) their relative school performance, and iii) their self-confidence (all of them 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:

$$logit_{perf}(TE) = \beta_0 + \beta_1 intell + \beta_2 av.grade \tag{10}$$

$$logit_{mot}(TE) = \beta_0 + \beta_3 homew.eff + \beta_4 subj.rank + \beta_5 self.conf \tag{11}$$

$$\begin{aligned} logit_{full}(TE) = \beta_0 &+ \beta_1 intell + \beta_2 av.grade \\ &+ \beta_3 homew.eff + \beta_4 subj.rank + \beta_5 self.conf, \end{aligned} \tag{12}$$

where (10) denotes the performance model, (11) the motivation model, and (12) the full model. The residuals of (10) to (12) are stored and will be used as predictors of students' probability to pass Abitur and to start academic studies, respectively (see Madon et al., 1997, for a similar operationalization of self-fulfilling prophecies). Following Gelman and Hill (2007, p. 97), the residuals r_i of logistic regressions are defined as $r_i = y_i - logit^{-1}(X_i\beta)$ – 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 (10) to (12). In this design, positive residuals indicate relative overestimations and negative residuals relative underestimations compared to the respective set of predictors in the logit models. For later analyses I 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).[19]

[19] An objection against this strategy may refer to the possibility of *private information*. More specifically, apart from 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

In terms of the distinction between *primary* and *secondary* effects of social inequality, in both the performance model and the full model, differences in the distribution of teachers' 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, in turn, are hypothesized to be the outcome of different teacher treatments.

3.3 Covariates

To keep track of the unobserved heterogeneity of my predictors (also see section 4.3), 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.

Parental educational attainment was measured by 13 categories reaching from lower secondary school without an apprenticeship up to a university degree. I categorized this variable into 1 'lower education'; 2 'middle education'; 3 'higher education' and 4 'degree' (see table A, appendix, for all summary statistics).

3.4 Models

In the empirical models I 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 also on a variable z_i that determines whether individual i will ever entry in the social situation or not.[20] 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^*, w' (on which the former 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*,

$$z_i^* = \mathbf{w}'\gamma_\mathbf{i} + u_i,$$

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 obscure), I tackle the implications of the former scenario in my robustness analyses in section 4.3.

[20] I follow the notation provided by Greene (2003, p. 782ff.).

and the equation of primary interest,

$$y_i = \mathbf{x_i'}\beta + \epsilon_i.$$

A typical example (which is taken from Greene, 2003, p. 782) is a model of female labor supply where the equation of primary interest is aimed to explain female respondents' wage y_i by a vector of predictors x_i (such as respondents' education and their job experience) with joint impact β on the outcome. However, a female respondent's wage is only observed if she is part of the labor market, i.e., if her number of labor hours z_i^* is higher than zero. The latter, in turn, could be determined by her marital status and home characteristics such as whether there are small children present.[21] 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. In consequence of this error correlation, conventional OLS regression merely considering the equation of primary interest yields inconsistent and inefficient estimates (also see Wooldridge, 2006, ch. 17.5).[22] This selection problem can be solved by a two step estimate 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 $\gamma^{'}$ are stored as *Inverse Mill's Ratios* λ_i which are computed as follows:

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

where ϕ is the $p$robability $d$ensity $f$unction (pdf) of the normal distribution with $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ for all real numbers x, and Φ is the $c$umulative $d$istribution $f$unction (cdf) with $\Phi(x) = \int_{-\inf}^{x}\phi(t)d(t)$ (which means to integrate over ϕ (Wooldridge, 2002, p. 458; Greene, 2003, p. 666).

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 since by controlling for λ as a metric instrumental variable for the exogenous determinants of the selection equation, also the problematic error correlation ρ – i.e. between u_i and ϵ_i – is canceled out.

This procedure can also be used if y_i is not completely unobserved for specific values of $z*_i$, but if x_i' is supposed to be an endogenous treatment that is affected by a vector of unobserved variables which are correlated with another vector of unobserved variables that influence y_i (Vella, 1993). In this case, the term "endogeneity bias" is common (e.g. Vella and Verbeek, 1999, p. 475). For instance, if students participate in a coaching program to improve their Scholastic Assessment Test (SAT) scores, and, apart from observed variables such as previous SAT scores and social backgrounds, also unobserved

[21] Note that gender-related child-raising practices implied in some econometric textbook examples do not necessarily correspond to the opinion maintained by the present study's author.

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

factors are supposed to affect their coaching program participation (e.g. 'grit'), and these unobserved factors are supposed to correlate with other unobserved factors that affect future SAT scores (e.g. 'moxie'), then the error terms between the coaching equation and the SAT equation would be correlated, and this endogeneity bias could be corrected by using the above-sketched Heckman methods (Briggs, 2004, p. 399).

In the present case, there is evidence to assume both sources of bias in the data. First, on the one hand, the distribution of the SEU indicators (x_i' in the above notation) is expected to vary strongly by parental social strata (w' in the above notation; see Becker, 2000, 2003) – which is the core idea of both the SEU model of educational transitions (Esser, 1999) and similar propositions since Boudon (1974). On the other hand, social backgrounds might affect both the definition and evaluation of the social situation, and thus also unobserved variables that influence the decision for or against a higher track of education. In econometric terminology, this is an example of endogeneity bias, and by regressing all SEU predictors on parental social class (selection equation) and including the IMRs of these estimates in the equation of interest, I should be able to control for the causal impact of unobserved 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 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).[23]

Regarding the self-fulfilling prophecy residuals, recall that one central shortcoming of the empirical literature of *Pygmalion* is 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. Moreover, and perhaps even more important, if unobserved variables affecting whether a student is over- or underestimated by her teacher are correlated with unobserved variables that determine her future educational and academic success – other variables that influence Δp_{ep} –, we should be able to reduce differences between the latter's 'real' value and our approximation in terms of SFP by applying the corresponding correction for endogeneity bias.

Second, evidently the propensities of the former students to start academic studies strongly depend on whether they successfully graduated from *Gymnasium* or not. Although for some particular school 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, a solution to the theoretical problem of conditional transition rates (Breen and Goldthorpe, 1997) should also account for the methodological problem of selection bias.

Concretely, I will estimate the following models: First, as mentioned, parents' expected benefit of higher education B, their subjective value of status decline $-SD$, the expected

[23] In the case at hand, this problem might be even stronger since the data at hand only contain records of *Gymnasium* students.

status decline p_{sd}, students' probability of successfully completing the chosen school track p_{ep}, the expected costs of education C, and the 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 the independent variables. The results of these probits will be stored as Inverse Mill's Ratios λ_{ij}. By subscripts i, j it is addressed 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 the first equation of primary interest which predicts the individual probabilities to pass Abitur:

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

Second, I will re-estimate (13) 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 (14), this model also includes *direct* controls for parental social class and educational backgrounds: [24]

$$logit_{UNI} = p_{ep}B + (1 - p_{ep})(-SD) - C + SFP + \lambda_{ij} + class + educ. \tag{14}$$

This procedure is summarized graphically in figure 2.

4 Results

4.1 Distribution of Variables

Dependent variables In figure 2a, 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 the chosen 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 2b captures the distribution of the time span until the former students began academic studies. Most of the students took up academic studies immediately after having passed Abitur, and a smaller number did so 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 began academic studies tremendously drops down. Thus, I choose this value as the cut-off for dichotomization of the second dependent variable.

[24] By introducing parental social class and education as additional covariates in (14), first, it is possible to test whether the parameters of the SEU-IEO model are an exhaustive specification regarding variation in parental cost-benefit considerations by social classes (see Stocké, 2007, for a case of remaining significant social background effects in the Breen-Goldthorpe model). Second, from the viewpoint of self-fulfilling prophecy research, we get an intuition whether the teacher treatment effect is mainly passed on students' social backgrounds as well (apart from achievement differences that have already been ruled out by residualizing teachers' expectations). In order to avoid identification problems of my Heckman model, I refrain from including parental social class and education also in (13) as the predictors therein have already been regressed on social class in the first-stage selection equations.

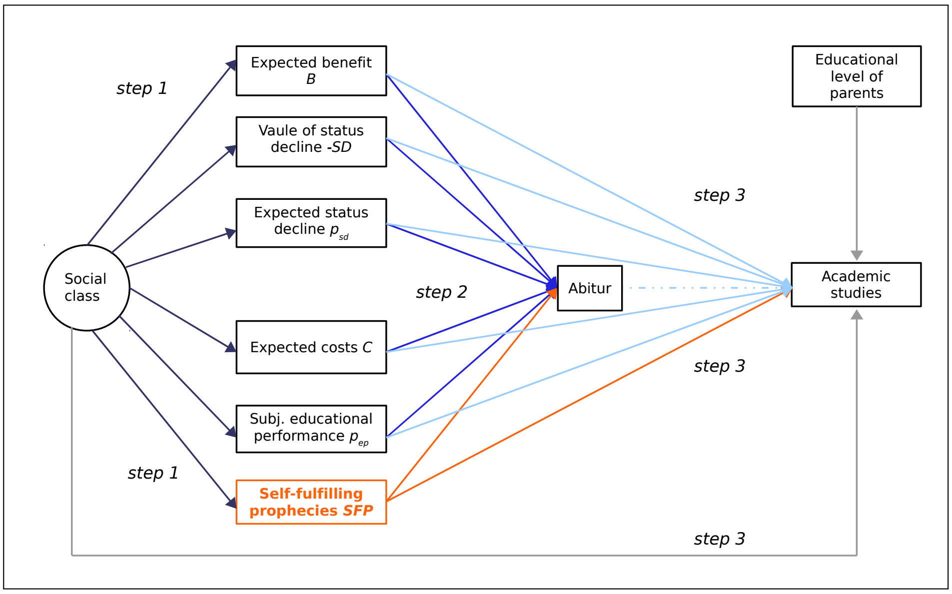


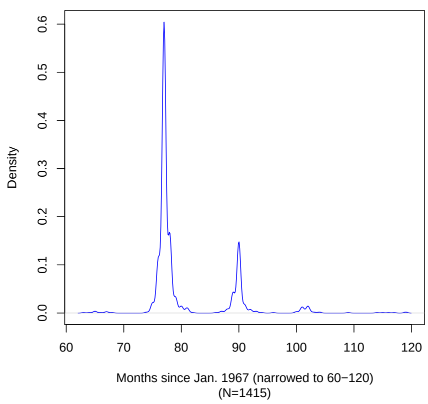
Figure 1: A modified model of Inequality in Educational Opportunities (modification of Becker, 2003, p. 7).

This procedure is also in line with more theoretical arguments: As Morgan (2002, p. 287f.) writes, "[t]he decision of whether or not to enter college immediately following high school is perhaps the most crucial determinant of alternative lifecourse transitions from adolescence to adulthood (...)" since delayed college entry "(...) yields different payoffs that result in alternative lifecourse outcomes". Hence, from panel data, it is of course possible to investigate university transitions at later points in time (see Hillmert and Jacob, 2010, for such an analysis based on the *German Life History Study*), but both related utility considerations and subsequent path dependencies may differ.

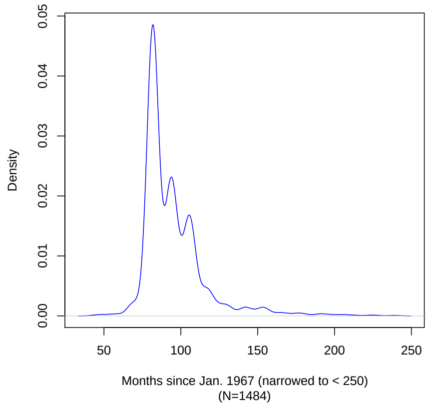
Main independent variable: teachers' evaluations Now I present the distribution of teachers' evaluations both numerically (figure 3a) and graphically (figure 3b). It can be noted that the amount of students who received a positive evaluation by their teacher (30.9%) is higher than the amount of students who received a negative one (25.4%) – but evidently most students did not obtain any evaluation at all (43.7%). As mentioned in section 3.2, for the following operationalization of self-fulfilling prophecies I will only focus on positive vs. negative teachers' evaluations.

Residuals of over- and underestimations Next I present the results of logit equations (9) to (11) that I use to extract the 'net' effects of self-fulfilling prophecies. Model 1 shows 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).[25] As regards the performance

[25] For these and all subsequent models, missing values have been deleted listwise (see Enders, 2010, for an elaborate discussion of the pros and cons of various missing value techniques under different missing data patterns).



(a) Distribution of educational success over time



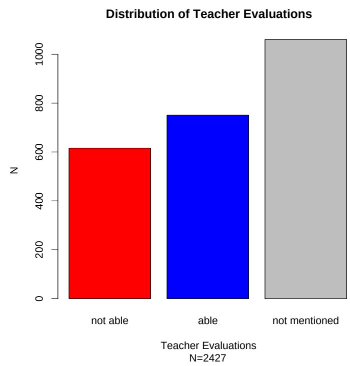
(b) Distribution of university transitions over time

Figure 2: Distribution of educational success and university transitions over time.

	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 teachers' evaluations



(b) Graphical distribution of teachers' evaluations

Figure 3: Distribution of teachers' evaluations: ability for academic studies.

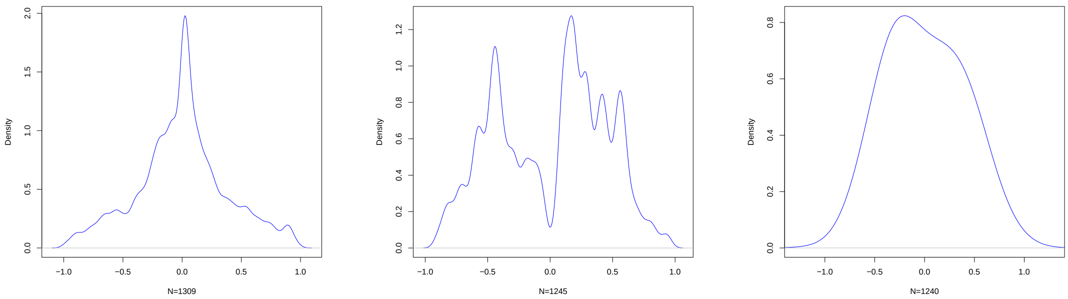
model (model 1), we can observe that both students' intelligence and their school grades significantly predict teachers' evaluations. The R^2 of this model is remarkably high. However, except the measure of students' relative school performance, in the motivation model, the z-values are much lower (self-confidence) 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 can conclude that for their teachers, students' performance is far more important than their motivation.

Table 1: Logistic regression of teachers' 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	1294	1287

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

As mentioned, I now store the residuals in order to use them as indicators of a 'net' effect of self-fulfilling prophecies. Figure 4 displays the distribution of the residuals from the three different logit models. Positive residuals indicate an overestimation relative to the predictors of the logit models, negative residuals a relative underestimation. In accordance with the predictive power of the performance model, the residuals in figure 4a mainly follow a normal distribution: Most students obtain an evaluation that is roughly on par 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 (figure 4b), we can note that students' motivation hardly suits to solely predict teachers' evaluations: Two local maxima can be found at 0.5 and -0.5, respectively – indicating that based on these background variables, the prediction of teachers' evaluations does not become more precise than simply by guessing. Finally, when we look at the distribution of the full-model residuals (figure 4c), we see that the curve gets slightly distorted, but is still very close to the normal distribution. As mentioned above, I dichotomized each residual for the following analyses.



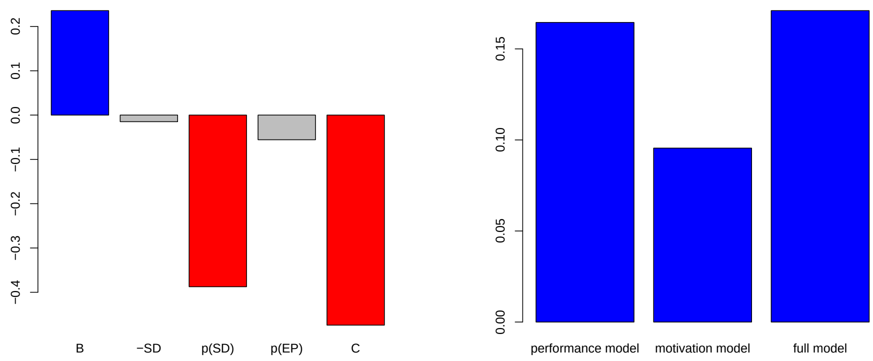
(a) Performance model (b) Motivation model (c) Full model

Figure 4: Distribution of residuals of teachers' evaluations on students' performance and motivation, and the combination of both.

As a validity check, I inspected the intercorrelations between the metric and dichotomized residuals and their predictors, respectively (table B, appendix), and I also computed kind-of reduced-form regression models of both educational success and university transitions wherein the residuals were joined by their predictors from table 1 (tables C and D, appendix). What follows from these results is that i) correlations between metric residuals and their predictors are negligible; ii) correlations between dichotomized residuals and their predictors are somewhat higher but remarkably low for the full model; iii) part of the explanatory power of the dichotomized residuals on students' educational success (cf. section 4.2) indeed *is* attributable to intelligence and average grade; but iv) both metric and dichotomized residuals still significantly predict students' educational success *when controlling* for the former variables; and v) what has been said for iii) and iv) also holds for the motivation model. In short, these validity analyses allow the conclusion that the residuals that have been computed in this section are not equal to their predictors whereby they had been estimated – but exert a distinct impact on students' educational success.

Bivariate probit estimates In this section I briefly discuss the results from the bivariate probit regressions of the SEU predictors of primary interest, B, p_{ep}, SD, p_{sd}, and C on parental social class, respectively (see figure 5a). 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 benefit 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 how parents judge the probability that their offspring might be impacted by status decline: While parents from lower social strata seem to be more concerned about the potential harmfulness of a lack of Abitur for their



(a) SEU predictors (b) SFP residuals

Figure 5: Standardized bivariate probit estimates of SEU predictors and self-fulfilling prophecy residuals on parental social class.

children's' later life, parents from the higher strata appear to be much more confident: Due to their higher resources, they feel that they might still be able to ensure their children getting ahead in life, even if the latter failed a final exam. With regard to $-SD$ and p_{ep}, no significant associations were found.

Figure 5b 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 overestimated by their teachers compared to their actual performance and motivation.

4.2 Multivariate Analyses

First I present the logistic regression estimates of students' probability of passing Abitur for both Esser's (1999) SEU predictors and the self-fulfilling prophecy residuals. I subsequently introduce the independent variables as well as the correction terms in the equations, so that I will present the following models: *Model 1a* contains the predictors for the additive interpretation of Esser's (1999) SEU model, B, $-SD$, p_{sd}, p_{ep} and C. In *model 2a* to *4a*, I 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 as shown in figure 6. *Models 1c* to *4c* repeat the same procedural method for the regressors that were constructed to measure Esser's (1999) theoretical concepts of "educational motivation", $B + p_{sd} \cdot SD$, and "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 suffered seriously from multicollinearity (the inter-correlations between the IMRs lie between an absolute value of .97 and .99), I summed up the IMR scores for all SEU predictors to one single

IMR score.[26]

Second, in *models 5a* to *8d* I 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 additive model interpretation are introduced in the model, the tables with the interaction terms are not discussed in depth here but are reported in the appendix (tables E and F).

Passing Abitur As table 2 shows, in the baseline model lacking both the self-fulfilling prophecy indicators and controls for potential sample selection bias, all SEU parameters except the perceived costs of education C have a significant impact on students' educational success in terms of passing Abitur. It is possible that costs do not come into play in this model because of mechanisms in line with the life course or selection hypothesis (Blossfeld and Shavit, 1993; Mare, 1980, 1993; Müller and Karle, 1993) which postulates that the effects of social inequality decrease in the course of students' education. However, students' chance of high school graduation still varies by the expected benefit of graduation, the expected amount of status decline and its expected impact, and by the subjective probability of educational success.

Interestingly, when introducing 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 the 95% level.[27] In the theoretical section I have argued that $p_{ep} = f(SFP)$ (8), and although, admittedly, I am 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 in terms 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 overestimated by their teachers with regard to their 10^{th} class academic performance compared to students who have been underestimated.

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

[26] The scale reliability of the IMR sum score is about Cronbach's $\alpha = .84$. Because in all cases, the inter-correlations between the single IMR scores are 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 the present case, a confirmatory factor analysis with factor loadings around .99 (not shown; available upon request) also strengthens the assumption of equal weights.

[27] A re-analysis of model 1a where observations with missing values on the self-fulfilling prophecy residuals are excluded shows that this drop-down in significance cannot be attributed to the reduced sample size in models 2a – 4a (cf. appendix, table G).

Table 2: Logistic regression of students' educational success on SEU predictors and self-fulfilling prophecy residuals

	Model 1a e^{b*sd}/z	*Model 2a* e^{b*sd}/z	*Model 3a* e^{b*sd}/z	*Model 4a* e^{b*sd}/z	*Model 1b* e^{b*sd}/z	*Model 2b* e^{b*sd}/z	*Model 3b* e^{b*sd}/z	*Model 4b* e^{b*sd}/z
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)	(3.14)	(1.86)	(2.15)	(2.67)
$-SD$	1.12*	1.21	1.23*	1.35**	1.31**	1.21	1.22	1.33**
	(2.05)	(1.82)	(1.96)	(3.05)	(2.87)	(1.77)	(1.91)	(2.91)
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)	(-0.76)	(-0.20)	(-0.20)	(-0.52)
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)	(4.77)	(2.28)	(2.22)	(3.56)
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)	(0.22)	(0.23)	(0.20)
r_{perf}		2.87***				2.85***		
		(10.47)				(10.37)		
r_{mot}			2.80***				2.78***	
			(10.26)				(10.16)	
r_{full}				1.84***				1.82***
				(6.37)				(6.23)
IMR_{sum}					1.16	0.67	0.65	0.83
					(1.39)	(-0.61)	(-0.65)	(-0.28)
IMR_{perf}						0.62		
						(-0.72)		
IMR_{mot}							0.60	
							(-0.76)	
IMR_{full}								0.74
								(-0.45)
Nagelkerke's R^2	0.06	0.36	0.35	0.21	0.12	0.35	0.35	0.21
N	1419	585	582	580	586	584	581	579

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}: subjective expected probability of educational success; 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).

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 increases when controlling for substantial over- and underestimations. As opposed to model 2a, the subjective expected benefit and the expected amount of status decline remain significant in *model 3a*.

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 teacher's 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 the SEU parameters in their additive empirical form 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} loses its significant impact on students' educational success. However, it is important to note that none of the self-fulfilling prophecy residuals are affected by sample selection correction. Since I indexed the Inverse Mills Ratios for the SEU predictors, I assume that this particular robustness is not an artifact 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 regression of students' transition propensities to university on both the SEU predictors and the self-fulfilling prophecy residuals is 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 probability of successfully completing the chosen school track, p_{ep}, have a positive impact on students' propensity to start academic studies. Both indicators for the expected status decline as well as expected costs of education remain insignificant.

As in table 2, the effects for B are partialed 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 diminish 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 within students' educational life course.[28] Neither parental social class nor education exert a significant effect, and the results do not differ when either one were removed from the

[28] 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 H (Appendix) indicates that the above trend also holds for the predicted probabilities of high school graduations and university transitions, respectively (see corrected model).

Table 3: Logistic regression of students' transitions to university on SEU predictors and self-fulfilling prophecy residuals

	Model 5a	*Model 6a*	*Model 7a*	*Model 8a*	*Model 5b*	*Model 6b*	*Model 7b*	*Model 8b*
	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z	e^{b*sd}/z
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	584	581	579	579	579	579	579

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}: subjective expected probability of educational success; 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).

models (not shown, available upon request).

Finally, in *models 5b - 8b*, I replicated models 5a to 8a with controls for sample selection. Therefore, I 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 the 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, I 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 do 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 of the 'conventional' SEU predictors suffered from sample selection bias, if students' propensity of university transitions is controlled for the selectivity of the sub-sample, also the estimates of teachers' over- and underestimations lack statistical significance.[29]

4.3 Sensitivity Analyses

One justifiable objection against the antecedent Heckit models (and, likewise, also against the models of Becker, 2003) addresses the predictors in the selection equations. In particular in the second-step selection equation (passing Abitur), the Inverse Mill's Ratios that had been stored from the first step might not perfectly 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 first-stage selection equation in the first step. Hence, the problem could arise that the third-step equation of interest might suffer from an identification problem because it includes a variable that also affects the instrument, i.e. the IMR control terms in the second-step selection equation.

To be precise, if exactly the same set of predictors is used in the selection equation as it is in the equation of primary interest, the model is still identified but only by functional form assumptions regarding the Inverse Mill's Ratios (Briggs, 2004). More concretely,

[29] Since the number of observations for the two model sets with and without sample selection equation are not equal, I repeated the analyses for models 5a-8a without the observations that 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 (appendix, table I). Furthermore, the validity analysis for the residuals indicated that the results might be sensitive against using metric residuals rather than their dichotomized counterparts. When I re-estimated all (uncorrected) models by using metric residuals, both the regression coefficients and z-values tended to be a bit lower than in case of using dichotomized residuals, but without losing significance (not shown, available upon request). A similar story holds when instead of the logit residuals, the *generalized probit residuals* (Gourieroux et al., 1987) are used – which are by construction uncorrelated with the predictor variables (Vella, 1998, p. 136) –; and even when the latter residuals are used to estimate a *bivariate probit* model (Holm and Jæ ger, 2011) rather than a 'manual' Heckit (not shown; available upon request).

the error term of the selection equation has to be normally distributed in order to ensure that the second equation is identified via the non-linearity of the Inverse Mill's Ratios obtained from the first-stage probit equation (Olsen, 1980; Duncan and Leigh, 1985). This kind of identification is sometimes referred to as *weak* identification (Vella, 1998, p. 135), because the linearity assumption could be violated (Winship and Mare, 1992, p. 341f.). For instance, Gronau (1974) observed that sample selection correction via the inclusion of Inverse Mill's Ratios does not work if a mutually exclusive and exhaustive set of dummy variables is included in both the selection equation and the equation of primary interest.

Therefore, it is recommended to *over*identify the model by including at least one additional regressor in the selection equation that is *not* part of the equation of primary interest (the *exclusion restriction*; also see Breen, 1996, p. 43f.). Ideally, the exclusion restriction should also be ensured to be uncorrelated with the outcome in the equation of primary interest in *theoretical* terms. Thus, even in the case of students' probability to pass Abitur as the equation of primary interest, parental social class as the selection-equation exclusion restriction might not perfectly fulfill a strong identification assumption.[30]

A strategy for situations wherein issues such as selection and unobserved heterogeneity might arise but good instruments that are not correlated with the outcome are not available[31] has been proposed by Buis (2010, 2011). The basic idea behind his suggestion is that unobserved variables u, which might affect both the main independent variable and the dependent variable over several transition points k, are captured by a weighted sum of random variables $\nu_k = \beta_{uk}u$ that in turn is approximated by a normal distribution.[32] To reflect a variety of scenarios regarding the distribution of this random variable, different values for the standard deviation of ν are assumed. If $sd(\nu) = 0$, 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 ν, the stronger the effect that is allowed for unobserved heterogeneity. In the present case, I will examine how the effects of all self-fulfilling prophecy residuals obtained before change with $sd(\nu) = 0, sd(\nu) = 0.5, sd(\nu) = 1$, and $sd(\nu) = 2$, respectively.

In more formal terms, the two-step Heckit estimates of section 4.2 are supplemented by a sequential logit model wherein the probability of university transitions is conditional

[30] Sartori (2003) proposes a bivariate probit estimator identifying a model that uses the same predictors for both the selection equation and the equation of primary interest by imposing the restriction that the error terms for both equations are equal. However, since we have in total six different selection equations to estimate, we would need a *multivariate* probit estimator with an equal error constraint for all six equations – which would involve extremely complex estimation procedures.

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

[32] Note that the notations in Buis (2010) and Buis (2011) differ. I here rely on the notation in Buis (2011).

on the sub-sample of those who have passed Abitur (Buis, 2010, 2011):

$$p_1 = \frac{exp(\beta_{01} + \beta_{11}SEU + \beta_{21}SFP)}{1 + exp(\beta_{01} + \beta_{11}SEU + \beta_{21}SFP)} \tag{15}$$

and

$$p_2 = \frac{exp(\beta_{02} + \beta_{12}SEU + \beta_{22}SFP)}{1 + exp(\beta_{02} + \beta_{12}SEU + \beta_{22}SFP)} \text{ if } pass_1 = 1. \tag{16}$$

Here β_{0k} is the intercept, β_{1k} is the regression coefficient for the vector of SEU predictors, and β_{2k} is the regression coefficient for the vector of self-fulfilling prophecy residuals for each transition equation. Note that equation (16) is only estimated for those 'at risk', i.e. for students having passed the first transition and successfully graduated from high school by obtaining the degree of Abitur, which is captured by the term $pass_1 = 1$.

Let $\Lambda(\cdot) = \frac{exp(\cdot)}{1+exp(\cdot)}$ denote the general functional form of the sequential logit model. Let further A, B, C refer to the three possible educational outcomes in the data at hand: leaving after high school without Abitur (A), passing Abitur but dropping out of the educational system (B), and passing Abitur and making the transition to university as defined above (C). If now the weighted sum of unobserved covariates at each transition, ν_k, is introduced, the expected probabilities of passing the two transitions averaged over ν_k read:

$$E_{\nu_1}(Pr[y \in \{B, C\}|SEU, SFP, \nu_1]) = E_\epsilon(\Lambda(\beta_{01} + \beta_{11} + \beta_{21} + \underbrace{\beta_{u1}u}_{\nu_1})) \tag{17}$$

$$E_{\nu_2}(Pr[y \in \{C\}|SEU, SFP, \nu_2, y \in \{B, C\}]) = E_\epsilon(\Lambda(\beta_{02} + \beta_{12} + \beta_{22} + \underbrace{\beta_{u2}u}_{\nu_2})) \tag{18}$$

The left-hand side of both equations can be understood as follows: In (17), the probability of passing Abitur without any statement about whether or not to leave further education thereafter is modeled for both SEU indicators and the SFP residuals. In (18), the probability of making the transition to university *conditional* on having passed Abitur is modeled for the same predictors as in (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 the 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 simulating different values for the standard deviation of the random variable, we can also approximate the potential impact of the unobserved covariate on the self-fulfilling prophecy residuals by assuming different values for the correlation ρ between the two variables.

I assume that by this technique I not only control for selectivity issues that may arise in a sequence of educational transitions, but I also tackle the objection of 'private information' that may be part of the teachers' evaluation heuristic without being reflected in the estimated self-fulfilling prophecy residuals. Because the scenarios, which are simulated for the weighted sum of unobserved covariates, also 'control' for a possible correlation with a specified independent variable, this may lead us to arrive at a better understanding about the direction that this private information may take.

Equations (17) and (18) have to be solved by means of numerical integration using maximum simulated likelihood (Train, 2003) which 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 shows the parameter estimates and their 95% confidence intervals of the three residuals for students' probability of passing Abitur for the scenario that the correlation between the unobserved covariate and each residual would be zero. On the horizontal axis, however, these estimates are plotted against different hypothesized 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 previously having passed Abitur. 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.[33] Starting 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 deflate the latter's estimates and likewise decrease their significance.

The same tendency holds for the parameter estimates of students' propensity of university transitions. Just like in the selection model of table 3, the results lack statistical significance for the case of $\rho = 0$ and $sd = 0$. If the correlation between ν 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 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 in both the estimates and their significance levels again. Yet, in that case the model would surely be impaired by multicollinearity, which would forestall an unambiguous interpretation (Farrar and Glauber, 1967).

In sum, we can conclude that in order to weaken the estimates and/or significance levels of the self-fulfilling prophecy residuals, the prerequisites for an unobserved variable 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 deflating its

[33] The coefficients and z-statistics of the variables of interest for these equations are listed in tables J (passing Abitur) and K (university transitions) in the appendix.

predictive power. Only if both conditions held up to a relatively large extent, the results would not be robust. Since I would expect this to be an issue of multicollinearity, I do not expect the latter phenomenon to thwart my 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, which goes beyond the conventional subjective-expected-utility (SEU) model of inequality in educational opportunities (IEO). My aim was first to develop a formal model, and second to test this model in order to predict students' probability to graduate from high school (*Abitur*) as well as their subsequent university transitions.

In the theoretical section, I started with summarizing the basic assumptions of the SEU-IEO model by Esser (1999, pp. 265-275). After a literature review of *Pygmalion* and self-fulfilling prophecy research (Rosenthal and Jacobson, 1968; Madon et al., 1997; Jussim and Harber, 2005), I brought in 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. I thus proposed an integration of self-fulfilling prophecies in the formal SEU-IEO model by Esser (1999) in terms of a teacher treatment effect on students' subjective expected probability of educational success.

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 my multivariate analyses that were based on the *C*ologne *Hi*gh *S*chool *P*anel (*CHiSP*), I 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, 2009a,b). This could be a corroboration of the life-course hypothesis (Mare, 1980, 1993; Müller and Karle, 1993) which 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 significantly predict both students' educational success in terms of passing Abitur and their university transition propensities. Thus, the tentative conclusion from these models would be that self-fulfilling prophecies have indeed distinct effects apart from 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.

As the variance of students' resources and preferences by social class, and the conditionality of their transition decisions brought up a selectivity problem, I 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 with respect to the prediction of their university transitions all self-fulfilling prophecy

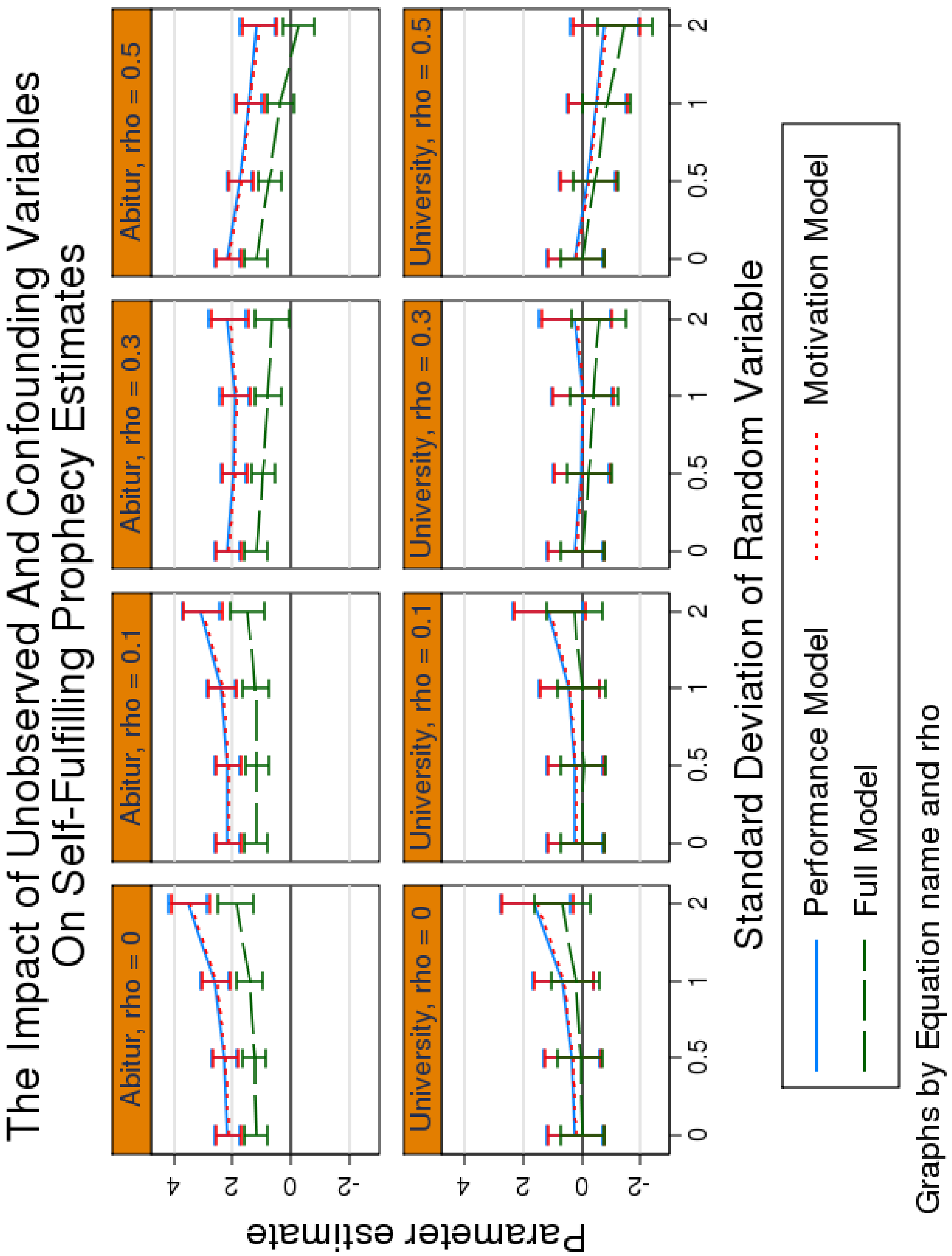


Figure 6: Sensitivity analyses of self-fulfilling prophecy residual estimates.

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' success in school, and that it does not influence their decision for or against starting academic studies.

Because of several methodological objections that could be raised against the quality of the instruments in the selection models, and in order 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, I 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 became apparent that only if relatively high values on both parameters are allowed for simultaneously, the residual estimates might not be robust. However, since this would go in line with the problem of multicollinearity, I do not expect my 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 of teaching in a particular class. Thus, future studies should also include potential covariates *apart from* the standard SEU predictors to ensure a better understanding of the social mechanisms behind the efficacy of self-fulfilling prophecies. This holds in particular if the empirical model *per se* is, as in the case of my 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

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 (2009a) 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.

Becker R and Hecken AE (2009b) Why are working-class children diverted from universities? An empirical assessment of the diversion thesis. *European Sociological Review* 25(2): 233–250.

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. In: Blossfeld HP and Shavit Y (eds.) *Persistent Inequality - Changing Educational Attainment in Thirteen Countries*, 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 (1996) *Regression Models: Censored, Sample Selected or Truncated Data.* Thousand Oaks, CA: Sage Publications.

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* 65(5): 754–772.

Briggs DC (2004) Causal inference and the Heckman model. *Journal of Educational and Behavioral Statistics* 29(4): 397–420.

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.

Buis ML (2011) The consequences of unobserved heterogeneity in a sequential logit model. *Research in Social Stratification and Mobility* 29(3): 247–262.

Card D (2001) Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica: Journal of the Econometric Society* 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.

Davis K and Moore WE (1945) Some principles of stratification. *American Sociological Review* 10(2): 242–249.

Duncan GM and Leigh DE (1985) The endogeneity of union status: an empirical test. *Journal of Labor Economics* 3(3): 385–402.

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.

Enders CK (2010) *Applied Missing Data Analysis*. New York: The Guilford Press.

Erikson RS and Jonsson JO (1996) Introduction: explaining class inequality in education: the Swedish test case. In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 1–63. Boulder, CO: Westview Press.

Esser H (1999) *Soziologie. Spezielle Grundlagen. Bd. 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.

Gill S and Reynolds AJ (1999) Educational expectations and school achievement of urban African American children. *Journal of School Psychology* 37(4): 403–424.

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.

Gourieroux C; Monfort A; Renault E et al. (1987) Generalised residuals. *Journal of Econometrics* 34(1-2): 5–32.

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

Gronau R (1974) Wage comparisons - a selectivity bias. *Journal of Political Economy* 82(6): 1119–1143.

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.

Holm A and Jæ ger MM (2011) Dealing with selection bias in educational transition models: The bivariate probit selection model. *Research in Social Stratification and Mobility* 29(3): 311–322.

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

Jackson M; Erikson RS; Goldthorpe JH et al. (2007) Primary and secondary effects in class differentials in educational attainment. *Acta Sociologica* 50(3): 211–229.

Jaeger MM and Holm A (2011) Relative Risk Aversion and Educational Decision Making: A Dynamic Model with Ability Updating and Heterogeneous RRA behavior. *Paper presented at the 2011 Spring Meeting of the ISA RC28, 13th-16th April 2011, University of Essex, UK* .
URL: *http://www.iser.essex.ac.uk/uploads/paper/document/38/rc28_jaeger_holm.pdf*

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 teachers' recommendations for secondary school track choice in Germany. *Paper presented at the 20th annual ESPE conference, 22 - 24 June, 2006, Verona* .

Jussim L (1986) Self-fulfilling prophecies: a theoretical and integrative review. *Psychological Review* 93(4): 429–445.

Jussim L (1989) Teacher expectations: self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology* 57(3): 469–480.

Jussim L; Eccles J and Madon S (1996) Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. In: Zanna MP (ed.) *Advances in Experimental Social Psychology*, 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.

Ladd HF and Walsh RP (2002) Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education review* 21(1): 1–17.

Lauen DL (2007) Contextual explanations of school choice. *Sociology of Education* 80(3): 179–209.

Lucas SR (2001) Effectively maintained Iinequality: 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.

Madon S; Willard J; Guyll M et al. (2006) Self-fulfilling prophecy effects of mothers' beliefs on children's alcohol use: accumulation, dissipation, and stability over time. *Journal of personality and social psychology* 90(6): 911–926.

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. In: Blossfeld HP and Shavit Y (eds.) *Persistent Inequality - Changing Educational Attainment in Thirteen Countries*, 351–376. Boulder, CO: Westview Press.

Mechtenberg L (2009) Cheap talk in the classroom: How biased grading at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies* 76(4): 1431–1459.

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

Meulemann H (1979) Klassenlage, Entscheidungsfeld und Bildungsaspirationen - Ein Versuch zur theoretischen Präzisierung und kausalen Erklärung von Zusammenhängen zwischen sozialer Struktur und individueller Lebensplanung. *Zeitschrift für Soziologie* 8(4): 391–414.

Mistry RS; White ES; Benner AD et al. (2009) A longitudinal study of the simultaneous influence of mothers' and teachers' educational expectations on low-income youth's academic achievement. *Journal of Youth and Adolescence* 38(6): 826–838.

Morgan SL (1998) Adolescent educational expectations. *Rationality and Society* 10(2): 131–162.

Morgan SL (2002) Modeling preparatory commitment and non-repeatable decisions. *Rationality and Society* 14(4): 387–429.

Muller C; Katz SR and Dance LJ (1999) Investing in teaching and learning: dynamics of the teacher-student relationship from each actor's perspective. *Urban Education* 34(3): 292–337.

Müller W and Karle W (1993) Social selection in educational systems in Europe. *European Sociological Review* 9(1): 1–23.

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

Nash R (2003) Inequality/difference in education: is a real explanation of primary and secondary effects possible? *The British Journal of Sociology* 54(4): 433–451.

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

Olsen RJ (1980) A least squares correction for selectivity bias. *Econometrica: Journal of the Econometric Society* 48(7): 1815–1820.

Parsons T (1959) The school class as a social system: some of its functions in American society. *Harvard Educational Review* 29(4): 297–318.

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. In: Cooper H and Hedges LV (eds.) *The handbook of research synthesis*, 301–321. New York: Sage.

Rivkin SG; Hanushek EA and Kain JF (2005) Teachers, schools, and academic achievement. *Econometrica: Journal of the Econometric Society* 73(2): 417–458.

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 DiMatteo MR (2001) Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology* 52: 59–82.

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* A25.

Sartori AE (2003) An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis* 11(2): 111–138.

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.

Sewell WH; Haller AO and Ohlendorf GW (1970) The educational and early occupational status attainment process: replication and revision. *American Sociological Review* 35(6): 1014–1027.

Sewell WH; Haller AO and Portes A (1969) The educational and early occupational attainment process. *American Sociological Review* 34(1): 82–92.

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. In: Hedström P and Swedberg R (eds.) *Social Mechanisms: An Analytical Approach to Social Theory*, 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.

Vella F (1993) A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review* 34(2): 441–457.

Vella F (1998) Estimating models with sample selection bias: a survey. *Journal of Human Resources* 33(1): 127–169.

Vella F and Verbeek M (1999) Estimating and interpreting models with endogenous treatment effects. *Journal of Business & Economic Statistics* 17(4): 473–478.

Wineburg SS (1987) The self-fulfillment of the self-fulfilling prophecy. *Educational Researcher* 16(9): 28–37.

Winship C and Mare RD (1992) Models for sample selection bias. *Annual Review of Sociology* 18: 327–350.

Wooldridge JM (2002) *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, Mass.: MIT Press.

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

6 Appendix

Table A: Descriptive statistics

	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: Correlation matrix of metric and dichotomized self-fulfilling prophecy residuals and their predictors

	teachers' evaluations	IST	average grade	motivation	self-concept	self-confidence	r(perf)	r(mot)	r(full)	r(perf.di)	r(mot.di)	r(full.di)
teachers' evaluations	1.00											
IST	0.32***	1.00										
average grade	-0.59***	-0.30***	1.00									
motivation	0.12***	-0.02	-0.12***	1.00								
self-concept	0.45***	0.20***	-0.52***	0.23***	1.00							
self-confidence	0.33***	0.11***	-0.38***	0.20***	0.62***	1.00						
r(perf)	0.75***	-0.00	0.02	0.07*	0.17***	0.12***	1.00					
r(mot)	0.88***	0.26***	-0.40***	-0.02	-0.00	0.00	0.76***	1.00				
r(full)	0.77***	0.00	0.00	0.00	0.00	-0.00	0.94***	0.87***	1.00			
r(perf.di)	1.00	0.32***	-0.59***	0.12***	0.45***	0.33***	0.75***	0.88***	0.77***	1.00		
r(mot.di)	1.00	0.32***	-0.59***	0.12***	0.45***	0.33***	0.75***	0.88***	0.77***	1.00	1.00	
r(full.di)	0.78***	0.08**	-0.18***	0.05	0.13***	0.11***	0.74***	0.80***	0.86***	0.78***	0.78***	1.00

All coefficients are Pearson correlations. Significance values: * ($p < .05$); ** ($p < .01$); *** ($p < .001$).

Table C: Logistic regression of students' educational success on self-fulfilling prophecy residuals and performance predictors

	Model r1a e^{b*sd}/z-value	*Model r1b* e^{b*sd}/z-value	*Model r1c* e^{b*sd}/z-value	*Model r1d* e^{b*sd}/z-value	*Model r1e* e^{b*sd}/z-value	*Model r1f* e^{b*sd}/z-value
r(perf)	1.93***	1.97***	1.99***			
	(8.08)	(8.35)	(8.96)			
r(perf.di)				3.76***	3.55***	2.47***
				(15.32)	(14.36)	(9.12)
IST		1.77***	1.42***		1.27**	1.20
		(6.92)	(3.56)		(2.60)	(1.81)
average grade			0.23***			0.43***
			(-12.15)			(-6.88)
Nagelkerke's R^2	0.12	0.20	0.46	0.40	0.41	0.47
N	813	813	813	813	813	813

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

Table D: Logistic regression of students' educational success on self-fulfilling prophecy residuals and motivation predictors

	Model r2a e^{b*sd}/z	*Model r2b* e^{b*sd}/z	*Model r2c* e^{b*sd}/z	*Model r2d* e^{b*sd}/z	*Model r2e* e^{b*sd}/z	*Model r2f* e^{b*sd}/z	*Model r2g* e^{b*sd}/z	*Model r2h* e^{b*sd}/z
r(mot)	2.68***	2.72***	2.73***	2.72***				
	(11.45)	(11.62)	(12.34)	(12.32)				
r(mot.di)					3.64***	3.59***	3.12***	3.10***
					(15.01)	(14.77)	(12.48)	(12.39)
motivation		1.47***	1.28**	1.27**		1.31**	1.23*	1.23*
		(4.79)	(2.76)	(2.59)		(3.07)	(2.33)	(2.23)
self-concept			2.60***	2.34***			1.53***	1.44**
			(9.68)	(7.39)			(4.26)	(3.16)
self-confidence				1.20				1.12
				(1.66)				(1.00)
Nagelkerke's R^2	0.24	0.27	0.42	0.42	0.39	0.40	0.42	0.42
N	804	804	804	804	804	804	804	804

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

Table E: Logistic regression of students' educational success on SEU interaction terms and self-fulfilling prophecy residuals

	Model 1c e^{b*sd}/z	*Model 2c* e^{b*sd}/z	*Model 3c* e^{b*sd}/z	*Model 4c* e^{b*sd}/z	*Model 1d* e^{b*sd}/z	*Model 2d* e^{b*sd}/z	*Model 3d* e^{b*sd}/z	*Model 4d* e^{b*sd}/z
$B + p_{sd} * SD$	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	585	582	580	586	584	581	579

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 F: Logistic regression of students' transitions to university on SEU interaction terms and self-fulfilling prophecy residuals with controls for sample selection

	Model 5c e^{b*sd}/z	*Model 6c* e^{b*sd}/z	*Model 7c* e^{b*sd}/z	*Model 8c* e^{b*sd}/z	*Model 5d* e^{b*sd}/z	*Model 6d* e^{b*sd}/z	*Model 7d* e^{b*sd}/z	*Model 8d* e^{b*sd}/z
$B + p_{sd} * SD$	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	584	581	579	579	579	579.00	579

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}: subjective expected probability of educational success; 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 G: Logistic regression of students' educational success on SEU predictors and self-fulfilling prophecy residuals (reduced sample size models)

	Model 1a e^{b*sd}/z-value	*Model 1a* (red. for perf.) e^{b*sd}/z-value	*Model 1a* (red. for mot.) e^{b*sd}/z-value	*Model 1a* (red. for full) e^{b*sd}/z-value
B	1.21***	1.33**	1.36***	1.34***
	(3.54)	(3.21)	(3.45)	(3.35)
-SD	1.12*	1.32**	1.32**	1.32**
	(2.05)	(2.97)	(3.00)	(2.94)
p_{sd}	0.88*	0.89	0.90	0.90
	(-2.18)	(-1.20)	(-1.17)	(-1.18)
p_{ep}	1.39***	1.66***	1.63***	1.63***
	(5.43)	(4.71)	(4.57)	(4.58)
C	0.96	0.96	0.97	0.96
	(-0.75)	(-0.40)	(-0.36)	(-0.39)
Nagelkerke's R^2	0.06	0.12	0.12	0.12
N	1419	585	582	580

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}: subjective expected probability of educational success.

Table H: 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 I: Logistic regression of students' transitions to university on SEU predictors and self-fulfilling prophecy residuals (reduced sample size models)

	Model 5e e^{b*sd}/z-value	*Model 6e* e^{b*sd}/z-value	*Model 7e* e^{b*sd}/z-value	*Model 8e* e^{b*sd}/z-value
B	1.23*	1.12	1.12	1.19
	(2.36)	(1.21)	(1.21)	(1.91)
-SD	1.16	1.07	1.07	1.18
	(1.69)	(0.67)	(0.67)	(1.78)
p_{sd}	1.02	1.06	1.06	1.03
	(0.16)	(0.60)	(0.60)	(0.32)
p_{ep}	1.74***	1.39*	1.39*	1.58***
	(4.41)	(2.46)	(2.46)	(3.57)
C	1.01	1.00	1.00	1.01
	(0.13)	(0.04)	(0.04)	(0.09)
r_{perf}		2.50***		
		(9.28)		
r_{mot}			2.50***	
			(9.28)	
r_{full}				1.65***
				(5.50)
social class	1.06	1.01	1.01	1.04
	(0.55)	(0.10)	(0.10)	(0.30)
education	1.14	1.07	1.07	1.11
	(1.16)	(0.55)	(0.55)	(0.87)
Nagelkerke's R^2	0.10	0.29	0.29	0.16
N	579	579	579	579

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}: subjective expected probability of educational success; r_{perf}: residuals (performance model); r_{mot}: residuals (motivation model); r_{full}: residuals (full model).

Table J: 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 K: 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

V. Does The Effect of Teachers' Expectations On Students' Educational Opportunities Decrease Over Educational Transitions?

A statistical matching approach

Various theoretical propositions suggest that social *i*nequalities in *e*ducational *o*pportunities (IEO) would decrease over students' educational transitions ('cross-grade patterns'; Lucas, 2009), but empirical analyses usually remain restricted to tests of student- or parent-level cost-benefit assessments. In this paper, I build on a subjective expected utility explanation of how teachers' expectations towards their students should affect the latter's transition propensities via their subjective expected probability of educational success, and I deduce hypotheses about decreasing teacher expectancy effects over students' educational course of life by referring to a mechanism of *Bayesian updating*.

Having outlined how my theoretical model suits the important *conditional independence assumption*, I test my hypotheses by means of a sequential logit model (with additional robustness analyses) based on an artificial student cohort that has been created via the technique of statistical matching (Rubin, 1986; D'Orazio et al., 2006). Results suggest that i) the predicted probabilities of educational transitions onto higher-level tracks are higher for students of whom their teachers hold more positive expectations, ii) the size of short-time expectancy effects decreases over educational transitions, and iii) short-term expectancy effects are larger than long-term expectancy effects.

1 Introduction

On the one hand, we find a variety of theoretical explanations such as the *L*ife *C*ourse *P*erspective (*LCP*; Müller and Karle, 1993), *M*aximally *M*aintained *I*nequality (*MMI*;

Raftery and Hout, 1993), *R*elative *R*isk *A*version (*RRA*; Breen and Goldthorpe, 1997) or *E*ffectively *M*aintained *I*nequality (*EMI*; Lucas, 2001) which all suggest that the effect of students' social backgrounds on their propensity to take the next higher node in the educational decision tree should decrease over educational transitions. This phenomenon which has been identified in many empirical studies (Mare, 1980, 1981; Smith and Cheung, 1986; Cobalti, 1990; Müller and Karle, 1993; Tolsma et al., 2007) was referred to represent *cross-grade patterns* (Lucas, 2009) – in contrast to the *cross-cohort patterns* (i.e. lower social background effects for later birth cohorts) that some of these theories either primarily or additionally aim to explain.

Although only one of the above-quoted theoretical propositions, namely RRA (Breen and Goldthorpe, 1997), was originally written down in a formal way, re-formalization of both MMI and EMI (Lucas, 2009) reveals that all three of them predominantly remain restricted to student- or parent-level cost-benefit analyses.

On the other hand, even in educational sociology it has already been criticized that standard economic approaches ignore the endogeneity of students' preferences in terms of an interaction with significant persons in the classroom (Lauen, 2007, p. 183) – although the latter circumstance had already been reflected in the early Wisconsin model of status attainment (Sewell et al., 1969, 1970; also see Morgan, 2002).

In the social-psychological literature, over four decades of *Pygmalion* (Rosenthal and Jacobson, 1968) and self-fulfilling prophecy research indicate that teachers' expectations can have considerable effects on students' educational achievement (for research overviews see Raudenbush, 1984; Jussim, 1986; Rosenthal, 1994; Jussim and Harber, 2005). This should, at least, result in path-dependency effects regarding students' educational transitions. Most notably, a great number of studies predominantly found teacher expectancy effects also to decrease over students' educational course of life (Rosenthal and Jacobson, 1968; Rist, 1970; West and Anderson, 1976; Frieze et al., 1991; Smith et al., 1999; Madon et al., 2006; Mistry et al., 2009; Hinnant et al., 2009; de Boer et al., 2010). These findings nicely coincide with the results for parental background effects obtained by educational sociologists.

Thus, the *theoretical* objective of this paper is first to build on an existing *s*ubjective *e*xpected *u*tility (henceforth referred to as *SEU*) explanation of how teachers' expectations towards their students should affect students' transition propensities via their subjective expected probabilities of educational success. Second, to deduce hypotheses about decreasing teacher expectancy effects over students' educational transitions, I point to a mechanism of *Bayesian updating* (Breen, 1999; Breen and García-Peñalosa, 2002; Morgan, 2005, ch. 5): Conditional on being over- or underestimated by their teacher, and conditional on having passed an educational transition or not, students either upwardly or downwardly update their beliefs according to the Bayesian rule. Based on a simplified decision tree, I elaborate on three different updating scenarios to clarify the argument.

From the perspective of conventional SEU explanations, the teacher side of classroom interaction is now explicitly considered. From the perspective of the current research status in teacher expectancy studies, I make use of a *repeated belief model* (Madon et al., 2006) that measures teachers' expectations at two distinct points in time; and I compute *residualized* indicators for teachers' expectations to construct over- and underestimations with regard to students' backgrounds such as their academic performance and their motivation (Madon et al., 1997, 2006; Hinnant et al., 2009; de Boer et al., 2010). The justification for this procedure is that unlike in experimental studies where perceivers' expectations can be manipulated by exposing them to biased information, in naturalistic (i.e. survey data) studies, only methodological techniques such as residualizing teachers' expectations can ensure them to be actually inaccurate to some extent.

The *empirical* contribution of the paper at hand is to discuss how the technique of statistical matching (Rubin, 1986) can be used to create artificial student cohorts for the analysis of social *i*nequalities in *e*ducational *o*pportunities (IEO) when suitable data is unavailable (also see Schubert and Becker, 2010). Since I do not dispose of one single data file comprising convincing measures of both teachers' expectations at different educational transition points and conventional SEU indicators, I use the distance *hot deck matching method* (D'Orazio et al., 2006) in order to combine one file containing teachers' expectations in primary school (Gesis-No. ZA893) with another one that contains teachers' expectations in 10^{th} grade (Gesis-No. ZA640). In the methodological section, I elaborate in more detail on what the important *conditional independence assumption (CIA)* requires from the theoretical distribution of the data and how my theoretical model can be assumed to suit this prerequisite.

The impact of teachers' expectations on students' IEO is then analyzed in the framework of a sequential logit model with additional robustness checks such as controls for unobserved heterogeneity (Buis, 2007, 2010, 2011), additional covariates and a comparison with analyses based on the British Cohort Study. Results suggest that i) the predicted probabilities of educational transitions onto higher-level tracks are higher for students of whom their teachers hold more positive expectations, and ii) that the strength of this effect decreases over students' educational life course.

2 Theory and Hypotheses

In the theoretical section, I first give a brief overview on both rational-choice and subjective expected utility explanations of inequality in educational opportunities (2.1). I then summarize the various theoretical propositions such as *LCP*, *MMI*, *RRA* and *EMI* that make statements about the waning coefficients pattern over students' educational transitions (2.2). Having described the general idea of *Pygmalion* or self-fulfilling prophecy research (2.3), I sketch my theoretical model regarding the effect of teachers' expectations on students' educational transitions (2.4). The theoretical section concludes by formulating falsifiable hypotheses deduced from the theoretical model that address how this effect can be assumed to decrease over educational transitions (2.5).

2.1 Inequality in Educational Opportunities: Rational Choice Approaches

The distinction between *primary* and *secondary* effects of social inequality (Boudon, 1974) is probably the most influential theoretical proposition to explain inequalities in educational opportunities at least in quantitatively-oriented educational sociology. The core idea is that *primary effects* are related to the impact of parental social backgrounds on their offspring's academic ability, while *secondary effects* refer to both structural and organizational conditions of the school as well as to the lower educational aspirations of the students themselves, or of their parents (also see Nash, 2003; Müller-Benedict, 2007). Differences in aspirations among social classes are then regarded to be the crucial factor explaining differences in educational opportunities even if differences in academic abilities were invariant among the social strata.

The Breen-Goldthorpe model Breen and Goldthorpe (1997) proposed a rational choice model to explain differences in educational transitions among social classes. Concretely, transitions are modeled as a function of subjective cost assessments c, the subjective likelihood of successfully completing the chosen school track π, and the utility that is attached to different educational outcomes. When α, β and γ denote the respective probabilities of having access to service class conditional on whether a student i) stayed and succeeded in school; ii) stayed and failed in school; or iii) left school, the probability of service-class children to remain in service class is 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

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

Breen and Goldthorpe (1997) show that $p_{is} > p_{iw}$ for any value of π less than one (p. 284). By content, this means that children from service class will be more willing to continue a high level of education than children from working class.

Apart from differences in resources, Breen and Goldthorpe (1997) highlight the conditional dependence of students' subjective probability of successfully completing the chosen school track, π_i, on their actual academic ability a_i: $\pi_i = g(a_i)$ (Breen and Goldthorpe, 1997, p. 285). This assumption has already proven useful for developing a formal model about the impact of self-fulfilling prophecies on students' educational opportunities (Becker, 2010a).

However, while the previous equations merely relate to class entry probabilities depending on a simple single-choice decision model, Breen and Goldthorpe (1997) acknowledge reality to be more complex. To be more precise, in most – if not all – cases of educational transition decisions, a *multiple* transition model approximates reality more closely. In this latter model, the number of parameters increases as there are now two transition choices that lead to different educational outcomes. First, we have two parameters of subjective expected educational success, π_1 and π_2, previous to each educational choice. Second, we have more parameters for students' probabilities of entering one of the three outcome classes: Let γ denote the probability of having access to service class after having left education at choice 1, β the corresponding probability after choosing to continue at choice 1 but failing thereafter, δ the probability after a successful way through education after choice 1 but after then deciding to leave at choice 2, ϵ the probability after a failure following the second transition, and Φ the probability of having access to service class after two successful educational transitions. Breen and Goldthorpe (1997) propose to set up the equation for the second transition decision first:

$$p_{is2} = \frac{\pi_{i2}\Phi + (1 - \pi_{i2})\epsilon}{\pi_{i2}\Phi + (1 - \pi_{i2})\epsilon + \delta} \tag{3}$$

Then, under the assumption that at the time when the decision for transition choice 1 is made the subjective expectations for transition 2 have already been shaped, Breen and Goldthorpe (1997) suggest to solve for α in equation 1 via backward induction:[1]

$$\alpha_i = q_{i2}(\pi_{i2}\Phi + (1 - \pi_{i2})\epsilon) + (1 - q_{i2})\delta, \tag{4}$$

where q_{i2} is a function of the subjective probabilities for the second transition, p_{i2} (Breen and Goldthorpe, 1997, p. 289). Thus, the expected probability of entering service class conditional on succeeding in education after the first transition choice is modeled as a function of students' expected probabilities after choice 2.

Esser's subjective expected utility model Esser (1999) takes up various ideas of both the Breen-Goldthorpe (1997) model and an earlier account by Erikson and Jonsson (1996) in introducing two additional terms that differentiate more precisely between the expected benefit of education, the expected amount of status decline and its impact on actual decisions. In this model, students' or parents' expected utility of continuing onto lower secondary school (A_n) is determined by

$$EU(A_n) = P_{sd}(-SD), \tag{5}$$

while the expected utility of continuing onto higher secondary school is determined by

$$EU(A_b) = P_{ep}B + (1 - P_{ep})P_{sd}(-SD) - C. \tag{6}$$

Here, SD is the expected amount of status decline with P_{sd} as its impact on parental decisions; B is the benefit of higher education (e.g. in terms of labor market prospects); P_{ep} is the subjective probability of successfully completing the chosen school track; and C are the expected costs of education (also see Becker, 2003; Pietsch and Stubbe, 2007).

[1] Note that the model of Bayesian belief updating proposed by Breen (1999) holds exactly the opposite view in that students' beliefs are supposed to change conditional on passing an educational transition. It is the latter position that I use to argue in favor of decreasing teacher expectancy effects in section 2.4.

By a series of linear transformations, Esser (1999) shows that $EU(A_b) > EU(A_n)$ if $B + P_{sd}SD > C/P_{ep}$. Adopting his terminology, this relation means by content that a higher level of education will be aspired if the educational motivation to continue somehow exceeds the underlying investment risk.

Although Esser's model doesn't make a statement about *conditional* educational transition decisions, its higher conceptual accuracy shall be adopted in the theoretical model of self-fulfilling prophecies sketched below.

2.2 The 'Waning Coefficients Pattern'

This section is aimed to give an overview of the theoretical approaches that try to explain why in numerous studies (Mare, 1980, 1981; Smith and Cheung, 1986; Cobalti, 1990; Müller and Karle, 1993; Tolsma et al., 2007), social background effects were observed to decrease in the course of students' educational transitions. Since most of these theories primarily intend to explain the phenomenon of waning coefficients over student cohorts rather than over educational transitions, the following allusions are far from exhaustive.

Life course perspective (LCP) In their study of the nine European nations participating in the CASMIN (*C*omparative *A*nalysis of *S*ocial *M*obility in *I*ndustrial *N*ations) project, Müller and Karle (1993) stressed that in order to explain the variance in social inequality between countries, scholars are advised to consider also the different class-specific survival patterns across transitions and nations. While France was found to be the most exclusive country, in Hungary the odds of completing secondary education are better than in most other countries (Müller and Karle, 1993, p. 9). However, for the purpose of the paper at hand, another one of their findings is much more important, which leads us to figure 1.

In this graph, Müller and Karle (1993) plot the estimates of a conditional logistic regression model of educational transitions on parental social class with varying success rates across both transition points and nations. The reference category of all lines are students from higher service-class families. Results show that at later transitions, class differentials tend to become smaller – resulting in a decrease in the relative risk of lower-class students not to pass the respective transition. This overall trend, inflections such as stronger social origin effects at transition T_2 compared to T_1 as well as other peculiar patterns are explained by the relation between the importance of different transition steps for the educational career and the variance of educational aspirations in the educational life courses among social strata.

Maximally maintained inequality (MMI) The crucial idea of the proposition of *m*aximally *m*aintained *i*nequality (*MMI*) can be summarized in the statement that "transition rates and inequality (as measured by odds ratios) remain constant unless forced to change by increasing enrollments" (Raftery and Hout, 1993, p. 42). By means of a cohort analysis of both the Irish Mobility Study and the Drumcondra Study of Educational Achievement, Raftery and Hout (1993) indeed find by trend decreasing social background effects at later educational transitions – but they also observe that this phenomenon is in turn sensitive to the progress of educational expansion: As enrollment rates to school increase, less margin for class differentials in entering secondary education is left (Raftery and Hout, 1993, 55f.). Thus, the difference between early and late transition background effects decreases for later cohorts as enrollment rates increase (figure 2).[2]

Relative risk aversion (RRA) Breen and Goldthorpe's explanation of *r*elative *r*isk *a*version (*RRA*) builds on their model of educational differentials due to varying aspirations and resources among social classes. Coming back to equations (3) and (4), Breen and Goldthorpe (1997) deduce that under the assumption that the parameters δ, ϵ, and ϕ do not vary by social classes, transition choices at higher levels will be less affected by social class than transition choices at lower levels: First, the situation of consecutive decisions will have reduced class differences in ability, and second, every higher-level transition decision is less risky in terms of suspected status decline than the preceding one. For instance, Breen and Goldthorpe (1997) assume that if secondary education had already been successfully finished, for working-class parents, a higher-level decision, say, for or against tertiary education would not be associated with the risk of demotion to the underclass anymore. As a result, class differentials concerning this decision may still persist but should be lessened compared to the preceding one(s) due to the lower perceived risk of downward mobility.

Effectively maintained inequality (EMI) The major advancement of Lucas' theorem of effectively maintained inequality is that it also considers that education may be tracked, and that students' selection into different school tracks may in turn vary by parental social backgrounds. That is, while MMI claims that socioeconomic differences will be canceled out when a certain level of schooling becomes universal, EMI posits that the socioeconomically advantaged will strive to secure qualitative differences (in terms of tracking) at the given quantitative level of schooling. In addition to this theoretical contribution, Lucas (2001) promoted IEO research also methodologically by demanding the analysis of predicted probabilities of educational transitions rather than of regression coefficients (or odds ratios) since the latter cannot reveal whether social background moves people over thresholds.[3] In an ordered probit model, Lucas (2001)

[2]The main difference between MMI and LCP is that "(...) LCP emphasizes that as children age they become more and more independent of parents, whereas MMI implies that adolescents' independence itself depends on the sociopolitical context and the support for particular levels of education" (Lucas, 2001, p. 1647).

[3]Apart from the statistical reason that distribution-insensitive predicted probabilities are robust against the criticism that was brought in by Cameron and Heckman (1998) and posits that the observed decline in odds ratios is only due to (unjustified) functional form assumptions of the underlying logit equations, Lucas (2001, 2009) also sees a substantive reason that speaks in favor of the analysis of predicted probabilities: Whereas odds ratios merely describe the association between predictor variables and their outcomes, only predicted probabilities consider "whether advantaged persons exceed pivotal thresholds to enter categorical positions that provide advantages" (Lucas, 2009, p. 490f.).

finds that both the parameter estimates of parental social backgrounds and students' predicted transition probabilities are lower in 12^{th} grade than in 11^{th} grade, but then rise again for college entry – which provides more support for MMI than for LCP. As regards EMI, Lucas observes that there are considerable differences between the social strata in their predicted probabilities of i) dropping out of school; ii) taking no course; iii) taking non-college preparation courses and iv) taking college preparation courses – which "is a far cry from the suggestion that social background effects decline to zero when a level of education becomes universal" (Lucas, 2001, p. 1678). Thus, according to this result, the data provides more support for EMI rather than for MMI.

Lucas' formal analysis of MMI, RRA and EMI Since LCP has already been rejected by Lucas (2001), in his formal analysis, Lucas (2009) only compares the implications of MMI, RRA and EMI. As he notes, only one of the three proposals, namely RRA, has already been written down in formal terms. Thus, Lucas' objective is first to provide a formal notation of both MMI and EMI, and second, to test which of the three proposals suffer from the logical threats *tautology*, *self-contradiction* and *evaluative infeasibility*.

In his formal analysis, Lucas (2009) shows that MMI is either contradictory or evaluatively infeasible. As regards RRA and EMI, both proposals stand the test of being non-tautological, non-contradictory and evaluatively feasible. As both RRA and EMI share a lot of common statements with yet a number of considerable differences, Lucas (2009) is unable to get to a final statement whether the two theories relate to each other in a complementary (or even nested?) manner. For the objective of the present study, the most important difference between RRA and EMI is that the former proposition explicitly links students' ability to both their subjective success expectations and educational achievement – while the latter one is agnostic on this relationship (Lucas, 2009, p. 502). As this is the crucial link the formal model of self-fulfilling prophecies presented by Becker (2010a) points to, RRA might be more suitable for deducing waning-coefficient hypotheses in this regard.

2.3 Self-Fulfilling Prophecy Research

Stemming on the idea of a self-fulfilling prophecy by Robert Merton (1948), Rosenthal and Jacobson (1968) showed in a (quasi-)experimental[4] setting that teachers' expectations can lead to their own fulfillment. In their famous *Pygmalion* study, an intelligence test was administered to elementary school children. Independently of children's actual test results, teachers were told the names of some randomly-selected students with the information that these students were likely to show a sudden intellectual spurt in the upcoming term. Interestingly, the students who had been labeled as 'late bloomers' scored significantly higher in a retest that was administered to the students one year after the initial test than the artificial control group. Thus, teachers' artificially-created expectations had actually become true.

[4]See Elashoff and Snow (1971) for a discussion of the experimental status of the initial *Pygmalion* study.

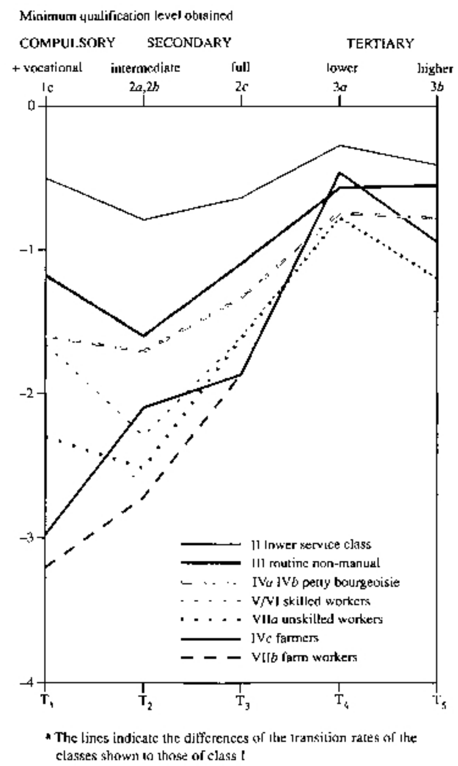


Figure 1: Parameter estimates for the effects of class origin on educational transitions in nine European countries (source: Müller and Karle, 1993, p. 13).

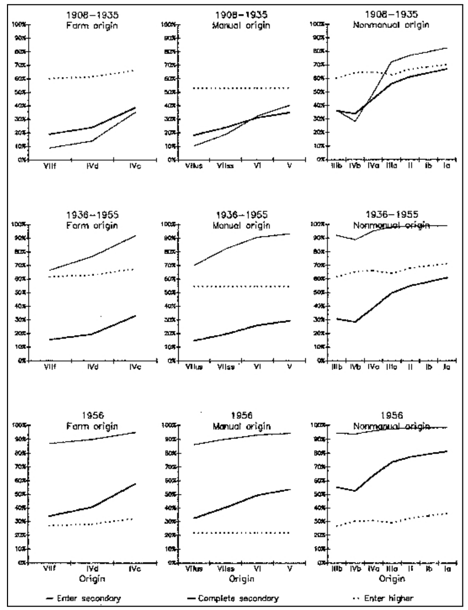


Figure 2: Successful transition to the next level of education (percentage), by origin and cohort: Republic of Ireland, 1921-75 (source: Raftery and Hout, 1993, p. 54).

Due to fundamental critics brought in by educational psychologists, Rosenthal became one of the pioneers in meta-analysis in order to defend his main idea of a self-fulfilling prophecy in the classroom. In a first meta-analytic review of the first 345 *Pygmalion* studies, Rosenthal and Rubin (1978) found an overall effect size of teachers' expectations on students' intelligence between $d = .14$ up to $d = 1.73$ and $r = .07$ up to $r = .65$. Later meta-analyses observed effect sizes between .11 (Raudenbush, 1984) and .16 (Smith, 1980).[5] Compared to the loads of basic *Pygmalion* and self-fulfilling prophecy studies, only few studies have bothered with searching for potential moderator effects[6] or with testing if effect sizes do accumulate, diminish, or remain stable over time. Concerning the latter question, following the review by Smith et al. (1999), until 1993, only four studies explicitly dealt with this question. First, Rosenthal and Jacobson (1968) themselves noted that the randomly-selected late bloomers gained more IQ points in the first year of the study than in the second year. Second, using table assignment as a criterion for identifying self-fulfilling prophecies, Rist (1970) noticed that differences between initial table assignment of the students in kindergarten and ability grouping by the teachers declined from first to second grade. Third, West and Anderson (1976) observed that the path coefficient relating teachers' expectations in students' freshman-year of high school to student achievement was higher for students' sophomore-year achievement than for students' senior-year achievement. Yet, fourth, Frieze et al. (1991) discovered that the (positive) relationship between facial attractiveness of MBA graduates and their salaries increased over time.[7]

Smith et al. (1999) themselves analyzed the effect of teacher perceptions in 6^{th} and 7^{th} grade on both students' final marks and standardized test scores in 7^{th}, 10^{th} and 12^{th} grade. As an advancement over the existing literature, they distinguished between *single-perceiver models* (comparing expectancy effects of the same teacher) and *multiple-perceiver models* (comparing expectancy effects of different teachers). Three out of seven different multiple-perceiver (i.e. teacher) analyses (all of them with marks as an outcome) give support for the dissipation hypothesis, another three analyses (two of them with marks, and one of them with standardized test scores as an outcome) give support for the stability hypothesis[8], and one analysis did not show self-fulfilling prophecy effects at all. The single-perceiver analysis also gave support for the dissipation hypotheses. Yet, Smith et al. (1999) noted that teachers' perceptions were able to predict student achievement up to seven years. Thus, although no support for the accumulation hypothesis was observed, self-fulfilling prophecies were considered to be long lasting.

In the last decade, about as much studies tested for potential changes in the self-fulfilling prophecy effect sizes over time than during the preceding three decades of

[5]See Rosenthal and DiMatteo (2001) for an explanation of common meta-analytical measures.

[6]See Jussim and Harber (2005) for a summary of research about moderators of self-fulfilling prophecies.

[7]However, note that Madon et al. (2006) consider neither the study by Frieze et al. (1991) nor the one by Rist (1970) to be 'real' self-fulfilling prophecy analyses since perceivers' beliefs have not been measured explicitly.

[8]In fact, the three analyses that Smith et al. (1999) named to give support for the stability hypothesis also showed declining coefficients for later achievement measures (which is at least suggestive in terms of the dissemination hypothesis), but the respective χ^2 difference tests lacked statistical significance.

Pygmalion and self-fulfilling prophecy research. In this context, the perhaps most profound study is the work by Madon et al. (2006). The main contribution of their paper is to test a *repeated belief model* measuring mothers' expectations about their children's alcohol consumption at different time points, and to examine potential changes in effect sizes over time. Whereas all above-cited studies only used 'gross' expectations to test for self-fulfilling prophecies, Madon et al. (2006) referred to the operationalization of Madon et al. (1997) who used *residualized* perceivers' expectations to construct over- and underestimations that are to a certain extent inaccurate.[9] They noted an accumulation of self-fulfilling prophecies for a combination of unfavorable expectations of mothers with regard to their children's alcohol consumption. However, Madon et al. (2006) did not treat the context of teachers' expectations on students' educational achievement.

Mistry et al. (2009) simultaneously tested the effects of both parents' and teachers' (unresidualized) expectations – both measured at two distinct time points – on students' achievement in the early school years. In several auto-regressive cross-lagged path models, Mistry et al. (2009) found stability of both adults' expectations and youths' test scores, but having controlled for teachers' expectations at t_2, no significant direct effects of teachers' expectations at t_1 on student achievement were found.

Hinnant et al. (2009) tested for effects of (residualized) first and third grade teachers' expectations on students' third and fifth grade reading and math abilities. Whereas they found no significant effect of teachers' expectations on students' reading abilities in neither third nor fifth grade, first grade teachers' expectations showed significant results on students' third and fifth grade math achievement, and third grade teachers' expectations also significantly predicted students' fifth grade math achievement.[10] As regards longitudinal comparisons, Hinnant et al. (2009) discovered that the effect of first grade teachers' expectations on students' third grade math achievement was larger than the effect of the former on students' fifth grade math achievement. The latter effect, in turn, was larger than the effect of third grade teachers' expectations on students' fifth grade math achievement.

And finally, de Boer et al. (2010) analyzed the effect of primary-school teachers' (residualized) recommendations regarding their students' secondary school track choice on the latter's performance up to five years later.[11] They found that in the first two years, teachers' expectancy bias partly dissipated but after that remained quite stable over

[9] In contrast to experimental studies such as the one by Rosenthal and Jacobson (1968) who directly exposed teachers to biased information and thus aroused biased expectations, naturalistic (i.e. survey data) studies have to ensure that perceivers' high or low expectations are actually inaccurate (also see Jussim, 1986). Madon et al. (1997) proposed to identify inaccurate expectations by regressing the respective indicator on a set of student achievement and motivation variables in order to use the residuals of these regressions as a measure of over- and underestimations with regard to the chosen set of predictor variables. Unfortunately, Madon et al. (1997) did not test for the stability of self-fulfilling prophecy effects over time.

[10] However, these results have to be interpreted with caution because Hinnant et al. (2009) additionally controlled for a teacher-based measure of students' social skills that could be correlated with teachers' expectations of students' achievements.

[11] As a methodological innovation, de Boer et al. (2010) applied a multilevel analysis in the first step and summed up the residuals of both levels for second-step self-fulfilling prophecy analysis.

time.

Overall, these results suggest that although self-fulfilling prophecies can be quite persistent over time, the strength of their effects tend to decrease over students' educational life course.

2.4 Self-Fulfilling Prophecies Over Educational Transitions: A Decision-Theoretic Approximation

Recently (Becker, 2010a), a proposition was presented of how self-fulfilling prophecies could be integrated into Esser's (1999) SEU model of educational transitions. I want to replicate the main idea of this model briefly before I try to deduce hypotheses that refer to potential *changes* of this relationship over time.

The main assumption of Becker (2010a) is that teachers' expectations affect students' future subjective expected probability of educational success via students' actual academic ability and a teacher treatment effect in terms of classroom praise, bilateral encouragement, and so on, that remains unspecified. Note that many self-fulfilling prophecy studies highlight that the crucial mechanism of self-fulfilling teacher expectations works via students' self-concept (Jussim, 1989; Gill and Reynolds, 1999; Muller et al., 1999; Mechtenberg, 2009; Mistry et al., 2009).

Becker (2010a) proposes to rewrite the Esser (1999) model as follows:

$$EU(A_b) > EU(A_n) \iff B + p_{sd}SD > \frac{C}{[\delta \cdot \tilde{p}_{ep_+} + (1-\delta) \cdot \tilde{p}_{ep_-}]} \tag{7}$$

– while $\delta = 0$ indicates that a student has been underestimated, and $\delta = 1$ that she has been overestimated by her teacher. $\tilde{p}_{ep_+}$ and $\tilde{p}_{ep_-}$ refer to her corresponding subjective expected probability of educational success, respectively, consisting of her initial value plus the gain (or loss) due to an over- (or under-)estimation. Then the denominator of the right-hand side of (7) should be larger on average for $\delta = 1$ than for $\delta = 0$, which allows the conclusion that also

$$EU_{\delta=1}(A_b) > EU_{\delta=0}(A_b). \tag{8}$$

This means that ceteris paribus, an overestimated student would expect a higher utility from choosing a higher-level school track than an underestimated one due to her higher subjective expected probability of educational success $\tilde{p}_{ep_+}$.

However, what is not addressed in this theoretical model is how $\tilde{p}_{ep_+}$ and $\tilde{p}_{ep_-}$ might change after *multiple* transitions. This can be explained following a framework of *Bayesian learning* as it has been proposed by Breen (1999), Breen and García-Peñalosa (2002), and Morgan (2005, ch. 5) for educational transition research. The core idea of this reasoning is that students update their beliefs while following the Bayesian rule reading

$$p(A|B) = \frac{p(A,B)}{p(B)}. \tag{9}$$

According to the Bayesian rule, the probability of A given that B is true equals the joint probability of A and B divided by the probability of B (since we already know B to be true).

Breen (1999) as well as Breen and García-Peñalosa (2002) proposed to model students' belief updating mechanisms – having passed a given transition – in a way of Bayesian learning that follows the above-shown rule. If θ' denotes a state wherein effort has a stronger impact on the probability of success than ability, and θ refers to a state wherein it would be just the opposite, and AH denotes a student's choice of a high educational career path, then the posterior belief of a student about θ' (i.e. the belief about θ' conditional on having succeeded in the high educational career path AH) reads (Breen and García-Peñalosa, 2002, p. 909):

$$Pr(\theta'|AH) = \frac{Pr(AH|\theta')Pr(\theta')}{Pr(AH|\theta')Pr(\theta') + Pr(AH|\theta)Pr(\theta)}. \tag{10}$$

Hence, a student's posterior belief about θ' is a function of her prior belief about θ' times her prior belief about her success in the higher academic track given that θ' is true (i.e. that effort is more important than ability) – divided by the term just described *plus* the prior belief that θ is true (i.e. ability is more important than effort) times the prior belief about her success in the higher academic track given that θ is true.

My aim here is not to present an exhaustive Bayesian analysis of the decision structure at hand – which might be worth a separate article –, but to use its main argument of updating for my line of reasoning. An important property of Bayesian updating is that Bayesian learning processes converge to a stationary belief (Breen and García-Peñalosa, 2002, p. 910), and below I discuss a decision-theoretical approximation that tries to capture this assumption for educational transitions and over- and underestimations, respectively.

A simplified decision tree of multiple over- and underestimations before and after multiple educational transitions is presented in figure 3. The tree is simplified as I only elaborate on students' subjective expected probability of educational success while holding all other SEU parameters constant. I assume that after each event – be it a teacher's over- and underestimation δ, or a student's educational transition τ – a student updates her subjective expected success probability according to Bayes' theorem.

While RRA (Breen and Goldthorpe, 1997) assumes that in the course of their educational transitions, students get more homogeneous regarding their academic ability, Breen (1999) postulates that due to Bayesian belief updating, students are *not required* to differ in their academic ability in order to produce differences in transition rates (e.g. between the social classes). In our case, by subsequent Bayesian updating after both obtaining a teacher's evaluation and making an educational transition, students may adopt their beliefs in a way that goes beyond initial ability differences – which also suits the model of self-fulfilling prophecies proposed by Becker (2010a).

How can we now deduce assumptions about the stability or volatility of self-fulfilling prophecy effects over time? First, from what was argued by Becker (2010a), $\pi_{1+} > \pi_{1-}$ – that is, the subjective expected probability of educational success of an overestimated student is larger than the corresponding parameter for an underestimated student. Second, following Breen (1999) and Breen and García-Peñalosa (2002), students who have successfully passed the first transition will update their subjective expected probability of educational success upwardly. Hence, $\pi_{2+}|\delta_1 = 1 > \pi_{1+}$, and $\pi_{2+}|\delta_1 = 0 > \pi_{1-}$.[12] Being over- and underestimated again, students will once more update their beliefs according to this teacher treatment.

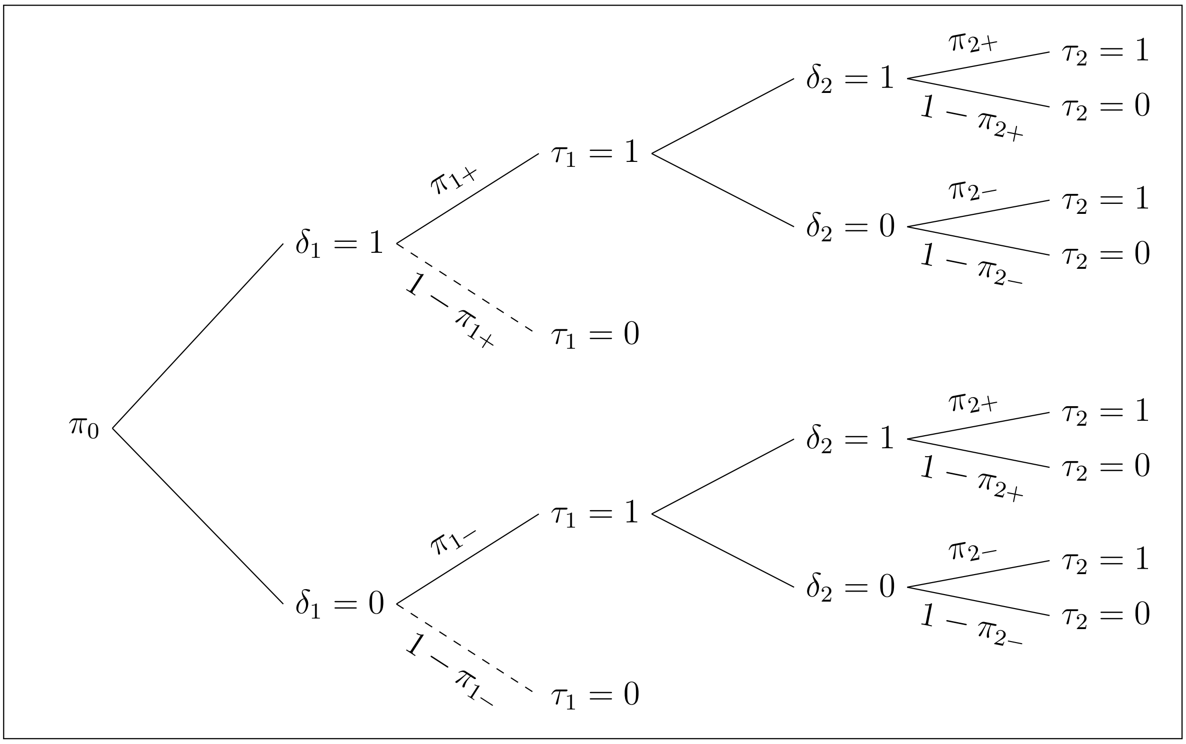


Figure 3: Decision tree.

As a starting point, let us assume that students have an initial subjective expected success probability of $\pi_0 = 0.5$. Let us further assume that in case of an overestimation, π increases by .1, and likewise, in case of an underestimation, π decreases also by .1. Moreover, assume that after a successful transition τ_1, the remaining students update their beliefs by $+.1$. The resulting subjective expected probability estimates before the second transition are listed in table 1 in the column that is labeled *Scenario 1.* Via an implied mechanism of Bayesian updating, a student who has been overestimated twice would finally show a subjective probability of .8; and a student who has been underestimated twice but nonetheless made the first transition would show a value of .4. Looking only at the final probabilities of students' who have been overestimated before the first transition ($\delta_1 = 1$) regardless of what happened later, we can average over

$$\pi|(\delta_1 = 1) = \frac{\pi_{2+}|(\delta_1 = 1, \delta_2 = 1) + \pi_{2-}|(\delta_1 = 1, \delta_2 = 1)}{2} = \frac{0.8 + 0.6}{2} = 0.7. \tag{11}$$

[12] Note that with the data at hand, students having not passed the first transition will drop out of the sample – which is why there is no need to bother with the beliefs in case of $\tau_1 = 0$.

Likewise, we can compute the average for students who have been *under*estimated *before* the first transition ($[\pi|\delta_1 = 0] = (0.6 + 0.4)/2 = 0.5$), the average for students who have been *over*estimated *after* the first transition ($[\pi|\delta_2 = 1] = (0.8 + 0.6)/2 = 0.7$) and the average for students who have been *under*estimated *after* the first transition ($[\pi|\delta_2 = 0] = (0.6 + 0.4)/2 = 0.5$).

If we now compute the difference between the average subjective probabilities of the over- and the underestimated students *before* the first transition, we can write

$$\Delta\pi(\delta_1) = \pi|(\delta_1 = 1) - \pi|(\delta_1 = 0) = 0.7 - 0.5 = 0.2; \tag{12}$$

while for the difference between average subjective probabilities of the over- and the underestimated students *after* the first transition similarly holds

$$\Delta\pi(\delta_2) = \pi|(\delta_2 = 1) - \pi|(\delta_2 = 0) = 0.7 - 0.5 = 0.2. \tag{13}$$

What becomes evident is that under the foregoing assumptions, differences in subjective probability estimates are equal for over- and underestimated students at both δ_1 and δ_2. Thus, if changes in subjective probability updating are equal after both a teacher's over- and underestimation and a student's educational transition, differences in subjective expected probabilities of educational success after having been over- and underestimated, respectively, are the same for the two possible time points of teachers' evaluations.

A similar phenomenon can be noted if we let the increase after educational transition $\tau = 1$ be .1, but let the change in π after being over- and underestimated constantly increase by only .05 (*scenario 2*): In that case, both $\Delta\pi(\delta_1)$ and $\Delta\pi(\delta_2) = .1$. Hence, differences in subjective expected probabilities after being over- and underestimated are not affected by differences in belief updating between an educational transition and teachers' evaluations.

However, let us finally take on the findings of self-fulfilling prophecy studies that teacher expectancy effects decrease over time. I argue that these decreasing effects are due to a mechanism of *changes* in belief updating; that is, the fact of being over- or underestimated affects a student's subjective expected educational success more strongly to an earlier point in time than to a later one. *Scenario 3* illustrates this assumption by setting the change in belief updating after the first teacher's evaluation δ_1 to .1, and after the second teacher's evaluation δ_2 to .05. Remarkably, under this scenario, $\Delta\pi(\delta_2) = 0.1 < \Delta\pi(\delta_1) = 0.2$ – which means that if teachers' evaluations affect students' subjective probability of educational success to a lesser extent at a later point in time than at an earlier one, then differences in students' *final* probability estimates are higher between the over- and underestimated ones at an earlier point in time than between the over- and underestimated at a later point in time.[13]

2.5 Hypotheses

Although Becker (2010a) already found self-fulfilling prophecy effects of teachers' expectations on students' educational success (in terms of Abitur) and their university

[13]For illustration purposes, the decision tree with the concrete values for scenario 3 is displayed in figure A (Appendix).

Table 1: Decision tree analysis under three scenarios

	Scenario 1 $\Delta(\pi\|\delta_1) = 0.1$; $\Delta(\pi\|\tau_1) = 0.1$; $\Delta(\pi\|\delta_2) = 0.1$	*Scenario 2* $\Delta(\pi\|\delta_1) = 0.05$; $\Delta(\pi\|\tau_1) = 0.1$; $\Delta(\pi\|\delta_2) = 0.05$	*Scenario 3* $\Delta(\pi\|\delta_1) = 0.1$; $\Delta(\pi\|\tau_1) = 0.1$; $\Delta(\pi\|\delta_2) = 0.05$
π_0	0.5	0.5	0.5
π_{1+}	0.6	0.55	0.6
π_{1-}	0.4	0.45	0.4
$\pi_{2+}\|(\delta_1 = 1, \tau_1 = 1)$	0.8	0.7	0.75
$\pi_{2-}\|(\delta_1 = 1, \tau_1 = 1)$	0.6	0.6	0.65
$\pi_{2+}\|(\delta_1 = 0, \tau_1 = 1)$	0.6	0.6	0.55
$\pi_{2-}\|(\delta_1 = 0, \tau_1 = 1)$	0.4	0.5	0.45
$\pi\|\delta_1 = 1$	0.7	0.65	0.7
$\pi\|\delta_1 = 0$	0.5	0.55	0.5
$\pi\|\delta_2 = 1$	0.7	0.65	0.65
$\pi\|\delta_2 = 0$	0.5	0.55	0.55
$\Delta\pi(\delta_1)$	0.2	0.2	0.2
$\Delta\pi(\delta_2)$	0.2	0.2	0.1

transitions based on analyses with the Cologne High School Panel (CHiSP), one objective of the present study is to replicate these results with the synthetic data at hand by testing whether self-fulfilling teacher expectancy effects also generalize with regard to the transition from primary to secondary school. Thus, the first hypothesis is that teachers' expectations affect students' educational transitions independently of conventional SEU predictors (H_1).

Second, as recent self-fulfilling prophecy research suggests, one can expect that although teachers' expectations might have enduring effects on students' educational life course, the magnitude of this relationship – in terms of predicted probabilities; see Lucas (2001, 2009) – should decrease over educational transitions (H_2). If we let $\Delta P(\delta_i, \tau_j)$ denote the difference in predicted probabilities P that is invoked by a teacher's over- or underestimation δ_i at a particular transition point τ_j (for a more illuminative description about the social mechanisms that are supposed to be captured in this parameter see Becker, 2010a), we could split up H_2 into two separately-falsifiable statements: On the one hand, it should simply hold that

$$\Delta P(\delta_1, \tau_1) > \Delta P(\delta_2, \tau_2) > \Delta P(\delta_3, \tau_3) \quad (H_{2a})$$

– which means that the effect of a teacher's expectation at t_1 on a student's first transition is larger than the effect of a teacher's expectation at t_2 on a student's second transition (and so forth). This is what I assume to follow from the mechanism of Bayesian updating, in that the effect of an over- or underestimation on the change in a student's subjective expected probability of educational success decreases between two time points of evaluations. On the other hand, one could expect that

$$\Delta P(\delta_1, \tau_2) < \Delta P(\delta_2, \tau_2); \Delta P(\delta_1, \tau_3) < \Delta P(\delta_2, \tau_2) \quad (H_{2b})$$

– which means that long-term effects of teachers' expectation are smaller than short-term effects. According to table 1, this is what follows from the previously-described mechanism for the final average probability differences between over- and underestimated students at two evaluation points.

In both cases, I assume that differences in subjective expected probabilities in educational success affect students' utility function as argued by Esser (1999) and thus also influence their actual transition decisions.

3 Operationalization

To the best of my knowledge, there exist no data that simultaneously collected both convincing measures of teachers' expectations at two different time points as well as tolerably accurate indicators for the SEU terms. Therefore, two different data sets have to be combined. The present paper will make use of the method of *statistical matching* (Rubin, 1986; D'Orazio et al., 2006) to create an artificial student cohort based on the statistical twins from two distinct data sources.[14] Before I elaborate on this method in more detail, a short description of the data and indicators available so far appears to be fruitful.

3.1 Data

The only data set known to me containing a measure of teachers' expectations in *secondary* school as well as reliable SEU indicators is the *C*ologne *Hi*gh *S*chool *P*anel (*CHiSP*; Gesis-No. ZA640). In the initial survey from 1969/70, N=3385 10^{th}-grade Gymnasium students in North Rhine-Westphalia were asked about issues such as their performance, interests and plans in school and about their social origin and their relationship to their parents. At about the same time, students' teachers (N=1701) and parents (N=2646) have also been surveyed.[15]

Taking this data set as a reference, we have to look for adequate primary-school data that were surveyed before *CHiSP* and provide the required information. A promising candidate is a data set named "Elternhaus und Bildungschancen" (*P*arental *H*ome and *E*ducational *O*pportunities, henceforth abbreviated as *PHEO*; Gesis-No. ZA893). In 1968, a population of N=1729 parents of 5^{th} grade students from Baden-Wuerttemberg whose children passed from primary to secondary school were surveyed. An additional

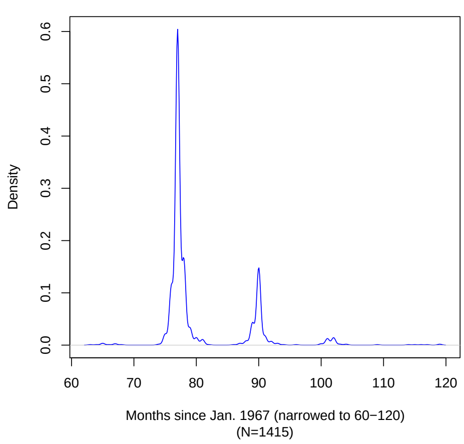
[14] In the empirical section, I will carry out additional robustness analyses based on the British Cohort Study data. But since some of the indicators therein are measured less precisely, I still see need for performing statistical matching of two files with better measures.

[15] Three re-surveys of the former students in 1984/85 (N=1987) 1996/97 (N=1596) and 2010 (N=1301) added private backgrounds and occupational careers of the former students' until age 30, age 43, and age 56, respectively.

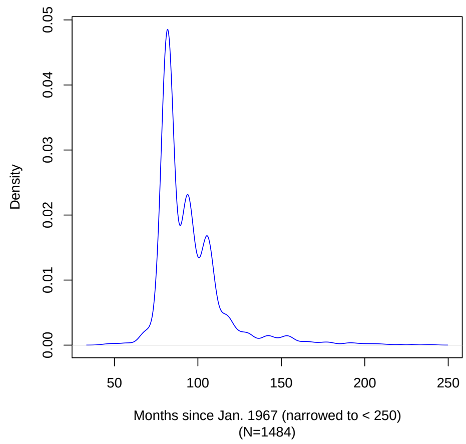
questionnaire collected rich information of primary school teachers (N=1426) including transition recommendations and other evaluations.

3.2 Variables

Dependent variables: educational transitions In line with the Mare (1980, 1981) model, the dependent variables of interest are students' educational transitions. From *PHEO*, we observe students' school choice after primary school. From *CHiSP*, we get information on the educational careers of a cohort of Gymnasium students. For the following analyses, a variable was constructed that takes 0 if a student did not pass to Gymnasium; 1 if a Gymnasium student quitted after 10^{th} class or later without passing Abitur; 2 if a Gymnasium student successfully passed Abitur but did not make the transition to university; and 3 if a Gymnasium student successfully passed Abitur and also made the transition to university. Since the re-surveys of the *CHiSP* asked if the former students *ever* passed the transition in question, I follow the approach proposed by Becker (2010a) in setting an ad-hoc cut-off value of 80 months after survey time in case of passing Abitur and of 106 months in case of transitions into tertiary education.[16] The distribution of high school graduations and university transitions over time are shown in figures 4a and 4b.



(a) Distribution of educational success over time



(b) Distribution of university transitions over time

Figure 4: Distribution of students' educational success and university transitions over time (original *CHiSP* data).

[16]For the *CHiSP* data, the zero point of counting has been backdated to January 1967. Thus, a cut-off value of 80 months includes all students who passed Abitur on the first try, but it does not exclude those who had to participate in immediate makeup exams. A cut-off value of 106 months includes all students who started academic studies within two years after high school graduation.

Independent variables The main independent variables of interest are primary school teachers' transition recommendations concerning their students' secondary school track choice (taken from *PHEO*), and secondary school teachers' 10^{th} class assessments whom of their students they consider to be able to complete academic studies successfully (taken from *CHiSP*). In *PHEO*, teachers could give a recommendation for *Hauptschule* (the lowest German secondary school track), for *Realschule* (the mid-level track) and for *Gymnasium*. I computed a new variable which was set to one if a teacher recommended Gymnasium and zero otherwise. In *CHiSP*, teachers could explicitly name students whom they considered to be able for academic studies, and of whom they considered to lack this prerequisite. I formed a new variable which was set to one if a student obtained a positive evaluation, and zero if she obtained a negative one. Students without an explicit evaluation were set to missing.[17] Leaning on the work by Madon et al. (1997, 2006), Hinnant et al. (2009) and de Boer et al. (2010), the appropriate setup for naturalistic, i.e. survey data self-fulfilling prophecy studies is to construct a measure of perceivers' over- and underestimations to ensure that their expectations are actually inaccurate. While the *CHiSP* data provides a great deal of promising regressors for teachers' expectations, the *PHEO* data is, unfortunately, less rich. For instance, in contrast to *CHiSP*, *PHEO* did not measure students' intelligence, which makes it more difficult to identify inaccurate teacher expectations. Instead, we have to rely on parents' assessments of students' school performance and motivation. To measure primary school students' performance in *PHEO*, I took the average of parents' statements about their children's grades in arithmetic, spelling, and literature (the only available grade information in that data). As a measure of their motivation, I controlled for parents' assessments of whether their children generally liked learning in school and of their children's estimated TV consumption time.

In *CHiSP*, we have more accurate measures of both students' school achievement and their motivation – but to keep the degree of 'accuracy' of the resulting over- and underestimations in both data sets comparable, I tried to use similar indicators also in *CHiSP*.[18] Concerning students' achievement, I also computed the average[19] of their grades in math and German classes; and with regard to their motivation, I controlled for their reported homework effort as well as for their self-confidence.

To construct inaccurate over- and underestimations, teachers' expectations at both time points were regressed on students' achievement[20] and their motivation. The residuals of these models were then stored as a new variable in order to use them as a predictor of students' educational transitions in the subsequent analyses. In total, I computed three different models that regressed teachers' expectations i) on performance

[17] For both a multinomial logistic regression and an ordinal structural equation model with teachers' evaluations as measured in *CHiSP* as an outcome that considers this 'missing' category, see Becker and Birkelbach (2011).

[18] For an operationalization of teachers' over- and underestimations based on a more precise set of students' achievement and motivation indicators taken from *CHiSP* data see Becker (2010a).

[19] If one of the grade variables was missing for a student, I computed the average of the remaining ones.

[20] Since in the German school system, *lower* grade points refer to a *higher* achievement, I have inverted the grades before computing the average.

indicators only; ii) on motivation indicators only; and iii) on both performance and motivation indicators together. For each of these three models, a separate residual variable was computed.[21]

Covariates According to my hypotheses, appropriate indicators of the SEU terms have to be controlled for. Once again, the precision of measurement is higher in the *CHiSP* data; hence, we should begin with a description of the respective variables therein in terms of an 'ideal state of affairs' to which the *PHEO* indicators have to measure up.

In *CHiSP*, students' expected benefit of eduction, B, was measured by the question if Abitur would be necessary for students to reach their aim in life (1 'yes, necessary'; 2 'useful, but not necessary'; and 3 'not important'). I recoded this variable into 0 'not important'; 1 'useful, but not necessary'; and 2 'yes, necessary'. The value of status decline, $-SD$, can be operationalized as parents' disappointment if their child would not pass Abitur (1 'not much'; 2 'little'; 3 'very disappointed'; 4 'would be the worst'). I categorized this variable into 0 'very disappointed/would be the worst'; and 1 'little/not much'. Concerning the expected probability of status decline, p_{sd}, parents' should assess the importance of good Abitur grades for their offspring's occupational success (1 'little'; 2 'not that much'; 3 'big'; 4 'very big'). This indicator was dichotomized into 0 'little/not that much' and 1 'big/very big'. The subjective probability of educational success, p_{ep}, was measured by parents' response if their child was able to complete Gymnasium (0 'probably not/don't know'; 1 'probably/definitely'). The expected costs of education, C, can be operationalized by parents' answer if they had to make financial sacrifices in order to offer higher education to their children (1 'no', 2 'little' and 3 'yes'). Once again I dichotomized the variable into 0 'no/little' and 1 'yes'.

In the *PHEO* data, the situation is a bit more difficult. A measure of the expected benefit of eduction, B, can be taken from several questions about the importance of the chosen school track and the aspired certificate for the intended occupational position and/or potential academic studies (0 'low'; 1 'average'; 2 'high').[22]

Instead of using Becker's (2003) operationalization of the value of status decline $-SD$ in terms of the difference between parental occupational prestige and the occupation anticipated for the offspring – which I consider *not* to suffice the prerequisite of measuring parental expectations in terms of subjective probabilities (Manski, 2004; also see Becker

[21] Tables A and B (appendix) show the estimates of the underlying logistic regressions.

[22] A *low* expected benefit was assigned to i) Hauptschule students whose parents had no idea about the utility of a General Certificate of Secondary Education (Realschulabschluss) or Abitur at all; ii) Hauptschule students whose parents wanted their child to take a job for which manual skills are most important; iii) Realschule students whose parents explicitly stated that Abitur would not be important for their offspring's future job; and iv) Gymnasium students whose parents only strove for a General Certificate of Secondary Education (Realschulabschluss). A *high* expected benefit was assigned to i) Realschule students whose parents aspired to move up to Gymnasium later on (surveyed by two different questions); ii) Gymnasium students whose parents said that Abitur is necessary for the occupation aspired for their offspring, or iii) in general improves her odds on the job market. Unfortunately, from the *PHEO* data it is not possible to generate a high estimated benefit for parents of Hauptschule students. Thus, the coefficient of B on the actual transition decisions may be slightly overestimated.

and Hecken, 2007, 2009a,b) –, I use two variables which indicate parents' concern that the chosen school track might be valueless. Students of both Haupt- and Realschule whose parents suspected that a General Certificate of Secondary Education might be useless were assigned a high amount of concern (1), while all others were assigned a low one (0). Similarly, Gymnasium students whose parents worried that Abitur might be useful only in the case of subsequent academic studies were assigned a high amount of concern (1), while all others were assigned a low one (0).

The expected impact of status decline, p_{sd}, can be measured more precisely, since there are questions about the general importance of school for achieving something in life in general and for occupational success in particular (0 'low'; 1 'high').[23]

The operationalization of the subjective expected probability of educational success, p_{ep}, is more problematic: While Becker (2010b) (and as far as I can see, this measure was also applied in Becker, 2003) specified to have used students' average grades in literature, spelling, and math – these indicators have to be discarded for this purpose in the present study since they have already been used for the identification of over- and underestimations. Instead, I use a combination of parents' indication whether their child had difficulties in primary school (0 'none'; 1 'a bit'; 2 'pretty') and their certainty about their decision of sending their offspring either to Realschule or Gymnasium. For parents who sent their children to Hauptschule, I additionally draw on an a question which asked whether this decision was due to low achievement of their offspring in some subjects.[24]

Fortunately, the expected costs of education C can be measured more straightforwardly: All parents were asked about the frequency of thoughts about getting along with their disposable money. Additionally, all parents of students who went to Hauptschule were asked about the expected costs of upper secondary school tracks as well as about the expected amount of these costs. Parents who said that they often worry if they had enough money until a month's end or how they should portion their money for the forthcoming week were assigned to experience high financial burdens (1), and zero otherwise. Furthermore, students of Hauptschule whose parents indicated that they would have expected much higher costs of education at a higher school track were also assigned a value of one.

According to Becker (2003), I additionally control for parental social class in terms of their occupational prestige and for parental educational attainment in order to capture potential social background effects on educational transitions that are not exhaustively modeled by the SEU indicators. In the *CHiSP* data, *parental social class* was measured

[23] I assume that the chosen indicators for p_{sd} capture more general attitudes towards the significance of school for success in life, while the selected variables for $-SD$ refer more to parental status concerns.

[24] I believe that this operationalization of p_{ep} in *PHEO* has two advantages: First, it ensures to use subjective measures of students' expected success that are not at conflict with the (more objective) indicators intended to be used for the identification of over- and underestimations. Second, although parents were surveyed at the beginning of secondary school, the temporal scope of the given questions aims at parents' expectations at primary school. Thus, potential difficulties in causality that could arise if one used secondary-school expectations to explain transitions from primary *to* secondary school should be attenuated.

by the Treiman occupational prestige scores of the head of each household; and *parental educational attainment* was measured by 13 categories reaching from lower secondary school without an apprenticeship up to a university degree. For harmonization purposes, I recoded this variable into four categories (0 'no General Certificate of Secondary Education'; 1 'General Certificate of Secondary Education but not more'; 2 'vocational or high school diploma'; 3 'finished tertiary education').

In the *PHEO* data, comparable scales of the head of household's occupation and education were formed.[25]

3.3 Method: Statistical Matching

Since I do not dispose of a single data file with convincing indicators for teachers' expectations measured before two distinct transition points, I use the method of statistical matching of the above-sketched *PHEO* and *CHiSP* data in order to create an artificial student cohort that suffices this prerequisite. In this subsection, I first allude to the *Conditional Independence Assumption* (CIA) as the crucial condition for performing statistical matching, and I defend how my theoretical model suits this condition. After that, I describe the matching algorithm that I used in practice.[26]

The Conditional Independence Assumption Let $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be a random variable with density $f(\mathbf{x}, \mathbf{y}, \mathbf{z}), \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, \mathbf{z} \in \mathcal{Z}$. Assume that A and B are two samples consisting of n_A and n_B independent and identically distributed observations generated from $f(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Let the units in A have $\mathbf{Z}$ missing, and the units in B have $\mathbf{Y}$ missing. Then the *Conditional Independence Assumption* (CIA) would mean an independence of $\mathbf{Y}$ and $\mathbf{Z}$ given $\mathbf{X}$.

For ease of understanding, start with the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ and $\mathbf{Z}$:

$$f(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z}) \tag{14}$$

Then the CIA just postulates that

$$f(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z}) = f(\mathbf{Y} \mid \mathbf{X}) \tag{15}$$

(cf D'Orazio et al., 2006, p. 13).

By content, this means that the distribution of $\mathbf{Y}$ given $\mathbf{X}$ and $\mathbf{Z}$ equals the distribution of $\mathbf{Y}$ given only $\mathbf{X}$ if $f(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z})$ does not depend on $\mathbf{Z}$. In that case, we can say

[25] While parental education as measured in *PHEO* could easily be harmonized to the variable that was computed in *CHiSP*, I had to assign Treiman scores (Ganzeboom and Treiman, 1996) to parents' answers about their field of occupation and their position therein by myself. Table C (appendix) lists the observed combinations of occupational *situs* and *status* in *PHEO* and with the assigned prestige scores.

[26] For all following formal issues concerning statistical matching see D'Orazio et al. (2006). Also cf. Rässler (2002).

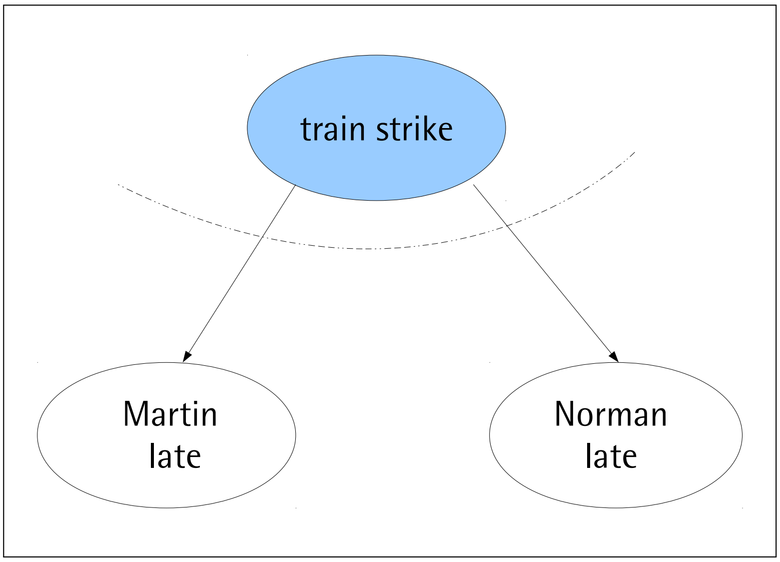


Figure 5: The Conditional Independence Assumption: An everyday-life example.

that **Y** is conditionally independent of **Z** given **X**. To provide a more illustrative example[27], consider the following situation: Imagine that two men, Norman and Martin, are inhabitants of the same city but live in two distant districts. Furthermore, they choose different traffic means to get to work (say Norman takes the train while Martin comes by car). Now X could denote the event "Norman comes late to work", and Z could refer to "Martin comes late to work". It might now appear that X and Z are completely independent from each other since the two actors are assumed to live on opposite sides of the city and do not use the same traffic means. However, there could be situations such as a train strike which accounts for both Norman coming late *and* also for heavy traffic volume on the streets leading also to Martin coming late (see figure 5).

A well-known problem is that the CIA can never be tested (let alone be proved). Thus, a cautious application of the CIA on the given research question has to reveal if the underlying theoretical model is in line with it.[28] In our case, **Y** and **Z** would be teachers' expectations in primary school and in 10^{th} class, respectively, as well as students' actual transition decisions. **X** would be the vector of SEU predictors and additional covariates. Do we find good reasons to defend the assumption that teachers' expectations at several points in time and students' transition decisions at several points in time do not influence each other, respectively, once the vector of SEU predictors is known? I hold the view that we do:

First, the crucial assumption of the SEU model (Esser, 1999) is that students' transition decisions can be explained by the vector of SEU predictors in the respective data set.

[27] I found this example on `http://www.eecs.qmul.ac.uk/~norman/BBNs/Independence_and_conditional_independence.htm`.

[28] Although the CIA is not less crucial for the application of propensity score matching when analyzing so-called *treatment* effects, it is often merely presupposed rather than justified on solid theoretical considerations (for instance in Caliendo and Kopeinig, 2008, p. 32).

That is, once we have controlled for parental utility considerations, two given educational transitions should be independent from each other according to theory.

Second, following the theoretical model outlined by Becker (2010a), and according to the above extension onto multiple evaluations that are separated from each other by a student's educational transition, a teacher's evaluation at t_1 should affect another teacher's evaluation at t_2 only via its impact on a student's *subsequent* subjective expected probability of successfully completing the chosen school track. The latter parameter is supposed to be a function of the student's *preceding* subjective expected probability of educational success, her actual academic performance, and an unspecified teacher treatment effect capturing classroom praise, bilateral encouragement, and so forth (Becker, 2010a, p. 10). Once we control for a suitable indicator of p_{ep} at both time points, intercorrelations between both teacher expectancy indicators should be substantively reduced.[29] This is why I would defend the assumption that for the given research question, the CIA should hold.[30]

Matching algorithms Basically, statisticians distinguish between parametric and non-parametric matching algorithms. Regarding *parametric* approaches, the most common one is *conditional mean matching* – though only working for normally distributed continuous variables. The idea of conditional mean matching is the prediction of two regression equations:

$$\bar{z}_a^A = \hat{\alpha}_Z + \hat{\beta}_{ZX} x_a^A \tag{16}$$

$$\bar{y}_b^B = \hat{\alpha}_Y + \hat{\beta}_{yX} x_b^B \tag{17}$$

Then each missing item in A is substituted with its expected value given the observed variables in B – and vice versa. For instance, assume that in file A, we observe respondents' education and their income, and in file B respondents' education and their work experience measured in years. Then we would regress work experience on education, and we would plug in the estimated mean value for a given value of education as obtained from B for all respondents with the same value of education in file A. However, as Little and Rubin (2002) note, this approach suffers from two serious drawbacks: First, some of the values of $\bar{z}_a^A$ and $\bar{y}_b^B$ may never be observed; and second, the variance estimators are not consistent. Therefore, I will directly turn to the explanation of the *non-parametric* approaches.

[29] Note that statistical matching researchers consider moderate relations between **Y** and **Z**, given **X** not as being consequential for statistical matching purposes (see e.g. Ingram et al., 2000, p. 5).

[30] Admittedly, in the present case, the validity of the CIA strongly hinges on the question if the SEU predictors in both data sets are reliable matching indicators. The demonstration of some measurement differences in both data sets might have led some readers to the conclusion that they are not. However, note that one elementary intention of the SEU model is to unveil the social mechanism providing an *understanding explanation* of how differences in educational transitions by social strata arise (Esser, 1999, p. 263). In other words, if the measurement of some SEU indicators in the *PHEO* data were too imprecise to predict students' transition decisions, then the predictive power of both parental education and their occupational prestige would disproportionally increase. Thus, by matching on the latter indicators, the CIA could still be maintained. Also see footnote 29.

Non-parametric matching approaches usually rely on some kind of hot deck method. In *random hot deck* matching, the idea is to randomly choose a donor record for each record in the recipient file after having grouped both files into homogeneous subsets (donation classes). For example, records in files A and B could first be grouped according to respondents' gender in order to ensure that women are always assigned to women, and men to men, after which they could be randomly matched.

Rank hot deck matching presumes that the units in both files are ranked separately according to the values of an ordinal matching variable (e.g. age categories). Then the youngest respondent of file B is assigned to the youngest respondent in file A, the second youngest observation in file B to the second youngest observation in file A, and so on.

And finally, the *distance hot deck* method matches each record in the recipient file with the closest record in the donor file according to a distance measure based on matching variables X. The basic decision rule is that

$$d_{ab} = |x_a^a - x_b^B| = min|x_a^A - x_b^B|; \tag{18}$$

that is, each observation in file B is matched to her closest mate in file A according to the chosen set of matching variables and a predefined distance function. Depending on the measurement level of the matching variables, the concrete distance function will be one of the corresponding measures such as the Euclidean metric, the Mahalanobis distance, the Gower dissimilarity coefficient, etc.

Another important distinction in this respect is the difference between *constrained* and *unconstrained* matching which touches the question whether a donor record shall be used more than once. Similar to an urn problem as it is well-known in elementary combinatorics, when performing statistical matching, each observation could either 'drop out' after it has been used once, or it can go back into the pool of records.

Furthermore, if the number of matching variables is large, additional donation classes can be predefined to decrease computational effort. That means that also when hot deck distance matching gets the researcher's vote, she can predefine 'fixed' classes such as respondents' gender or their education wherein distances between all observations of the two files are minimized separately.

Data harmonization and matching variables For the following analyses, *PHEO* and *CHiSP* shall be matched using an unconstrained distance hot deck technique with donation classes. Distance hot deck matching is an approach that is superior to parametric methods for the reason that

> "(1) imputations tend to be realistic since they are based on values observed elsewhere; (2) imputations will not be outside the range of possible values; and (3) it is not necessary to define an explicit model for the distribution of the missing values" (Siddique and Belin, 2008, p. 84).

As D'Orazio et al. (2006) note, several issues have to be considered when preparing data in order to perform statistical matching. In particular, the two data sources can be *biased* or *inconsistent* (also see van der Laan, 2000). Sources are *biased* if their samples are characterized by different reference periods or drawn from different populations. In our case, the given research question demands that the data refer to different periods since the aim is to construct an artificial student cohort. However, the reader may have noted that the respective survey times do not exactly correspond to the requested years. An additional problem arises with respect to the two samples' population: While *PHEO* consists of primary school parents in Baden-Wuerttemberg, *CHiSP* is composed of secondary school parents in North Rhine-Westphalia. Institutional factors in the two German Federal Lands may lead to difficulties in comparability, thus the following analyses have to be interpreted cautiously.

The issue of *inconsistency* refers to the harmonization of matching variables. As mentioned before, at least for some variables, measurement differences between the two sources are undeniable. Therefore, it has to be resolved if this does not result in loss in predictive power.

As D'Orazio et al. (2006) suggest, it has to be clarified that the matching variables $\mathbf{X}$ are associated with the variables of interest, $\mathbf{Y}$ and $\mathbf{Z}$. In case of categorical outcomes, this can be accomplished by means of a classification tree (Breiman et al., 1984). The idea of a classification tree – and of its metric counterpart, the regression tree – is to explain the variation in the response variable by repeatedly splitting the data into more homogeneous subgroups. A deviance measure based on χ^2 statistics is used to find the split that maximizes the dissimilarity among the resulting subjects (also see Hansen et al., 1996; De'ath and Fabricius, 2000). In contrast to conventional discriminant analysis, classification and regression trees do not follow a simultaneous partitioning logic but pursue a hierarchical approach, wherein each subgroup – graphically represented by a tree node – will be split up according to a distinct criterion (for a review of more elaborate classification techniques see Prasad et al., 2006).

Figures 6 shows the categorization trees for students educational transitions.[31] We can see that in case of the decision for or against the transition to Gymnasium, the expected benefit of education B is the most dominant predictor, while the value of status decline $-SD$ comes second, followed by highest parental educational degree and the subjective expected probability of successfully completing the chosen school track, p_{ep}.[32]

Contrarily, in the case of passing Abitur, p_{ep} is most prevalent while B comes second. Hence, as a preliminary result we can note that at a later educational transition i) less predictors suffice to explain its variance (which corresponds to the waning coefficients pattern); ii) subjective probability expectations are more important than considerations of educational benefit or status decline; and iii) the impact of parental social backgrounds dissipates (at least in a categorization tree model). However, since all of the selected indicators in the classification tree of the *PHEO* data are still *bivariately* associated with passing Abitur in the underlying χ^2 tests (not shown, available upon request), I

[31] I used the `tree` package in R (Ripley, 2010) for both computation and graphical display of the categorization tree. The minimum number of observations to be included in either child node was set to 5 (which is also the default for the `tree` package).

[32] As mentioned before, the explanatory power of parental education in *PHEO* could be a hint that some of its SEU indicators might suffer from a loss in precision due to measurement difficulties. Therefore, it appears fruitful to ground the matching procedure also on parental education.

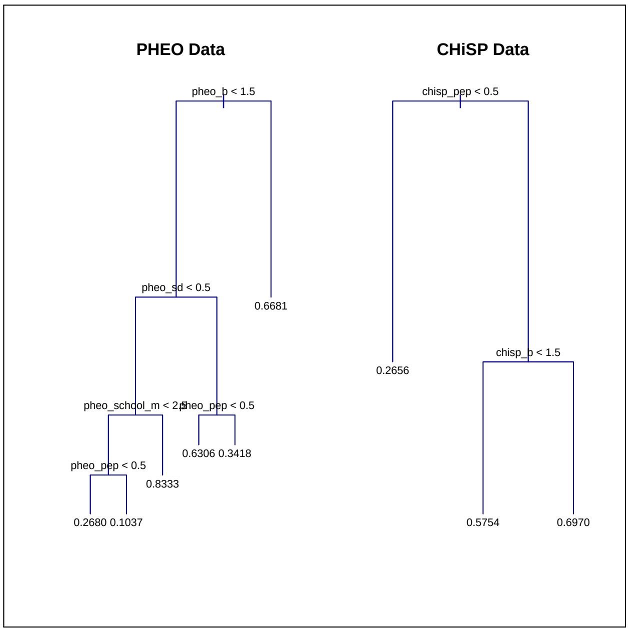


Figure 6: Classification trees of matching variables (outcome: educational transitions).

nonetheless consider these indicators to be reliable matching variables.

Usually, it is recommended to choose the smaller file as the recipient in order to avoid problems in affecting the distribution of the imputed variables in the final synthetic file (D'Orazio et al., 2006, p. 35). However, in our case, starting with the $N = 1729$ observations of *PHEO* would equal a considerable loss of the $N = 3385$ observations in *CHiSP* – which would then lead to a worrisome attrition in cell frequencies at the higher nodes of the transition tree. Therefore, I choose the unorthodox way of matching the smaller file, *PHEO*, to the larger file, *CHiSP*, to prevent loss in statistical power due to small cell frequencies.

From the categorization tree analysis, we identified B, $-SD$ and p_{ep} as the most promising matching indicators. To decrease computational effort, and in line with both theoretical considerations and the results from the classification tree, I use parental educational attainment as donation classes. Since the matching variables showed both binary and ordered categorical levels of measurement, the Gower distance function was applied.[33] I used the `StatMatch` package in *R* (D'Orazio, 2009) to create the synthetic file.

Table 2 shows the distribution of the relevant variables in *PHEO* data that were matched to *CHiSP* before and after the matching procedure. We can see that there are only minor differences between the two sets of variables except the distribution of parental education. The latter case is a result from matching *PHEO* to *CHiSP* (and not

[33] The general formula for the Gower distance function is $d_{ab} = \frac{1}{P}\sum_{p=1}^{P} c_p d_{abp}$ where a and b are sets of records with dimension P, and c_p is a scaling factor for the p^{th} variable. The idea is to use suitable distances for variables with different measurement levels, such as $c_p = 1$ for binary variables and $c_p = 1/R_p$ (i.e. the range) for continuous and categorical variables (D'Orazio et al., 2006, p. 216).

the other way round) since *CHiSP* – surveyed in the highest German secondary school track – is more selective in parental composition than *PHEO*. But since the crucial SEU indicators are quite comparable between the two files, this allows the conclusion that the matching procedure led to reasonably consistent results.[34]

3.4 Models: Sequential Logit Analyses With Unobserved Heterogeneity

As a first step, I apply conventional logistic regression analysis to estimate the log-odds of primary school children to move on to Gymnasium. For this step, all variables stem from *PHEO*. As a second step, I use sequential logit modeling (Buis, 2007, 2010, 2011) to estimate the conditional log-odds of passing Abitur and university transitions, respectively. These first two steps cannot be unified to a comprehensive sequential logit model since in the first case, I need explanatory variables from primary school, and in the second case, secondary-school predictors are required.

Yet, to test H_{2b}, in a third step I analyze changes of primary-school predictors over all subsequent educational transitions. Here, I can thoroughly analyze a comprehensive sequential transition model starting with the transition from primary to secondary school and ending with the transition to university.

The implied process of this artificial but comprehensive transition model is illustrated graphically in figure 7 (also see Buis, 2011, p. 3). Note that each capital letter does not denote an educational transition but relates to a particular educational *outcome*.

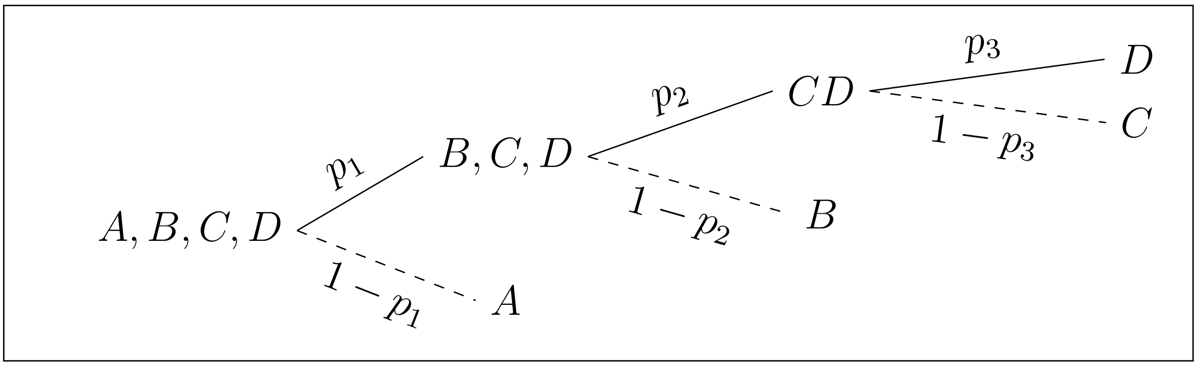


Figure 7: Educational transition process.

[34]A comparison of the regression coefficients for secondary school choice as a binary outcome between the variables of the original *PHEO* data and those that were matched to *CHiSP* will give further insight in the reliability of the synthetic file. A complete list of summary statistics for all independent variables in the synthetic file is given in table D (appendix).

Table 2: Descriptive results of donor data (*PHEO*) variables

	count	mean	sd	min	max
B	1729	1.05	0.51	0.00	2.00
-SD	1721	0.21	0.41	0.00	1.00
p(sd)	1652	0.32	0.47	0.00	1.00
C	1708	0.50	0.50	0.00	1.00
p(ep)	1723	0.29	0.45	0.00	1.00
par. occ. pres.	1650	45.31	12.68	18.00	72.00
par. occ. pres. (dichot.)	1729	0.63	0.48	0.00	1.00
par. educ.	1622	0.31	0.74	0.00	3.00
transition recommendation	1737	0.37	0.48	0.00	1.00
residual (perf.)	1549	0.00	0.44	-0.86	1.12
residual (mot.)	1453	0.00	0.48	-0.48	0.83
residual (full)	1393	-0.00	0.44	-0.86	1.13
residual (perf, dichot).	1549	0.45	0.50	0.00	1.00
residual (mot., dichot.)	1453	0.38	0.48	0.00	1.00
residual (full, dichot.)	1393	0.44	0.50	0.00	1.00

(a) Before matching.

	count	mean	sd	min	max
B	3374	1.13	0.52	0.00	2.00
-SD	3358	0.24	0.43	0.00	1.00
p(sd)	3244	0.32	0.46	0.00	1.00
C	3339	0.48	0.50	0.00	1.00
p(ep)	3367	0.26	0.44	0.00	1.00
par. occ. pres.	3213	49.34	14.12	18.00	72.00
par. occ. pres. (dichot.)	3374	0.70	0.46	0.00	1.00
par. educ.	3374	1.14	1.23	0.00	3.00
transition recommendation	3133	0.49	0.50	0.00	1.00
residual (perf.)	3001	0.09	0.43	-0.86	1.12
residual (mot.)	2748	0.10	0.49	-0.48	0.83
residual (full)	2634	0.07	0.43	-0.86	1.13
residual (perf, dichot.)	3001	0.56	0.50	0.00	1.00
residual (mot., dichot.)	2748	0.47	0.50	0.00	1.00
residual (full, dichot.)	2634	0.54	0.50	0.00	1.00

(b) After matching.

Assume that A refers to the outcome of not making the transition from primary school to Gymnasium, B to making the transition to Gymnasium without passing Abitur on the first try, C to passing Abitur on the first try without starting academic studies after 106 months, and D to starting academic studies after 106 months after having successfully passed Abitur on the first try. The important assumption of the sequential logit model is that the probability of each educational outcome is evaluated against the probability of all other *possible* educational outcomes, conditional on preceding transitions. For example, the probability of passing Abitur on the first try, p_2 (with the possible option of a subsequent university transition within 106 months after high school graduation), is evaluated against the probability of *not* passing Abitur on the first try, $1 - p_2$. But of course, it is taken into account that this step is conditional on having made the transition from primary school to Gymnasium.

Regarding modeling strategy, I first introduce the SEU predictors (Esser, 1999) in their *additive* interpretation and then subsequently add each residual term (from the performance model, the motivation model, and the full model, respectively) separately. I present the SEU terms in their additive interpretation since it might be of interest to assess the temporal stability of each single indicator – which would not be possible in the constructed terms for what Esser (1999) calls *educational motivation* and *investment risk*. As Becker (2010a) observed, results for the self-fulfilling prophecy residual terms obtained from the full *CHiSP* sample are stable for both the additive and the multiplicative reading of the SEU model. The same holds for analyses based on the synthetic file (not shown, available upon request).

Hence, for the artificial student cohort, the following models will be estimated (also see Buis, 2011, p. 3):

$$p_1 = Pr(y_1 = 1|\mathbf{SEU}, SFP_i, \mathbf{SES}) = \frac{exp(\beta_{01} + \beta_{11}\mathbf{SEU} + \beta_{21}SFP_i + \beta_{31}\mathbf{SES})}{1 + exp(\beta_{01} + \beta_{11}\mathbf{SEU} + \beta_{21}SFP_i + \beta_{31}\mathbf{SES})} \tag{19}$$

$$p_2 = Pr(y_2 = 1|\mathbf{SEU}, SFP_i, \mathbf{SES}, y_1 = 1) = \frac{exp(\beta_{02} + \beta_{12}\mathbf{SEU} + \beta_{22}SFP_i + \beta_{32}\mathbf{SES})}{1 + exp(\beta_{02} + \beta_{12}\mathbf{SEU} + \beta_{22}SFP_i + \beta_{32}\mathbf{SES})} \tag{20}$$

$$p_3 = Pr(y_3 = 1|\mathbf{SEU}, SFP_i, \mathbf{SES}, y_2 = 1) = \frac{exp(\beta_{03} + \beta_{13}\mathbf{SEU} + \beta_{23}SFP_i + \beta_{33}\mathbf{SES})}{1 + exp(\beta_{03} + \beta_{13}\mathbf{SEU} + \beta_{23}SFP_i + \beta_{33}\mathbf{SES})} \tag{21}$$

– where **SEU** refers to the vector of SEU parameters, SFP_i to the three self-fulfilling prophecy residuals (introduced separately), and **SES** to controls for parental education and their occupational prestige. The conditions $y_1 = 1$ in (20) and $y_2 = 1$ in (21) state that the respective equation is estimated only for the students "at risk", i.e. those who made the preceding educational transition.

4 Results

After showing the univariate distribution of students' educational transitions in the synthetic file, I present the log-odds of the sequential logit models, and I also display the respective predicted probabilities. The analysis section ends with a robustness analysis of the current results.

4.1 Distribution of Educational Transitions

Table 3 displays the distribution of students' educational transitions in the matched data file. 1677 observations or an amount of 63.16 % of the synthetic sample of N=2655 observations did not move on to Gymnasium, and 444 observations or a share of 16.72 % left Gymnasium without Abitur. 56 observations or an amount of 2.11 % passed Abitur but did not make the transition to university, and 478 observations or a share of 18 % both passed Abitur and moved on to university.

Table 3: Distribution of educational transitions (synthetic data)

	n	pct	cumpct
no Gymnasium	1677	63.16	63.16
Gymnasium without Abitur	444	16.72	79.89
Abitur without university transition	56	2.11	82.00
Abitur with university transition	478	18.00	100.00
Total	2655	100.00	

4.2 Multivariate Analyses: Sequential Logit Modeling

In this section, I first discuss the results from sequential logistic regression analysis in terms of conventional regression coefficients (log-odds). Following Lucas (2001, 2009), I then elaborate on the predicted probabilities.

Regression coefficients Table 4 compares the logistic regression estimates of primary school students' transitions to secondary school in both the original *PHEO* data and the synthetic file. Differences in significance can be noted for all three estimated coefficients for the expected impact of status decline p_{sd} with controls for the self-fulfilling prophecy residuals, and for one coefficient for perceived costs of education C. Thus, we have to keep in mind that our synthetic file will potentially overestimate the effect of p_{sd} on students' educational transitions. The latter point might particularly be relevant for C the t-statistic whereof ($t = -1.80$ in model 3a) is not far from the critical value of -1.96.[35] Notably, however, the coefficients for the residuals that were used to identify teachers' over- and underestimations are quite similar, and all of them significant, which gives support for H_1.[36]

[35] To ensure that these differences are not due to differences in sample size between original *PHEO* and the synthetic file, I took 100 random samples of size N=794 from the synthetic file and re-estimated model 4b. The results were virtually identical to the ones reported in table 5b (not shown; available upon request).

[36] A previous study (Becker, 2010a) found that the self-fulfilling prophecy residuals are insensitive to both their measurement level (metric vs. dichotomized) and the link function of the underlying generalized linear model wherefrom they are obtained (logistic vs. probit regression). The same holds for the residuals generated from the synthetic data at hand (not shown, available upon request).

Table 5 presents the results of a sequential logit model of students' educational transitions. Models *5a* to *8a* refer to students' probability of passing Abitur, and models *9a* to *12a* to students' university transitions. In model 5a, only the SEU predictors went into the equation. We can see that students' expected educational benefit B and their subjective expected probability of successfully completing Gymnasium p_{ep} are positively associated with their probability of passing Abitur. These results dissipate in models *6a* and *7a* when the residuals from the performance model and the motivation model are introduced, respectively. From both models we can learn that the more a teacher overestimates her student compared to her actual performance or motivation, the higher the probability that she successfully passes Abitur. This result still holds when teachers' evaluations are regarded net of both students' performance and their motivation altogether (model *8a*). The only differences we encounter compared to the two preceding models are that i) both effect size and z-value are a bit lower, and ii) students' subjective expected probability of successfully completing Gymnasium p_{ep} still significantly predicts their probability of passing Abitur. Thus, the results from these models still provide support for H_1.

Turning to models *9a* to *12a* that list the estimates for the prediction of students' university transitions, we do not find any parameter that would be significantly associated with that outcome. Hence, the tentative conclusion up to now would be that both the impact of the SEU predictors and the effect of self-fulfilling prophecies decrease over educational transitions.

In models *5b-8b* (table 6), students' expected costs C is the only primary school SEU indicator that significantly predicts students' probability of passing Abitur (which only holds if the self-fulfilling prophecy residuals are part of the equation). However, we do not find any significant effect of either one of the self-fulfilling prophecy residuals that capture teachers' primary school over- and underestimations – which is in line with hypothesis H_{2a}.

Likewise, in models *9b-10b*, neither one of the SEU predictors nor the self-fulfilling prophecy residuals are significantly associated with students' propensities of university transitions. This result would give support to hypothesis H_{2b}, but note that we still have to compare the predicted probabilities before drawing final conclusions.[37]

[37] For all self-fulfilling prophecy residuals and all except two SEU terms, the results of tables 5a - 6 are robust against controls for both parental education and occupational prestige (see Appendix, tables Fa - G).

Table 4: Logit model of secondary school transitions

	Model 1a log-odds/z	*Model 2a* log-odds/z	*Model 3a* log-odds/z	*Model 4a* log-odds/z
B	2.43***	2.32***	2.02***	2.37***
	(10.70)	(8.48)	(6.13)	(8.13)
-SD	1.37***	1.48***	1.40***	1.52***
	(6.93)	(5.89)	(4.65)	(5.61)
p(sd)	0.02	0.07	-0.18	-0.02
	(0.13)	(0.33)	(-0.68)	(-0.11)
C	-0.49**	-0.53**	-0.45	-0.50*
	(-3.14)	(-2.66)	(-1.80)	(-2.33)
p(ep)	-1.58***	-1.40***	-0.98***	-1.57***
	(-9.21)	(-6.59)	(-3.74)	(-6.85)
residuals (performance model)		2.67***		
		(12.90)		
residuals (motivation model)			4.09***	
			(14.63)	
residuals (full model)				2.68***
				(12.10)
Nagelkerke's R^2	0.41	0.60	0.74	0.61
N	995	889	830	794

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.

(a) Original *PHEO* data.

	Model 1b log-odds/z	*Model 2b* log-odds/z	*Model 3b* log-odds/z	*Model 4b* log-odds/z
B	1.55***	1.40***	1.28***	1.32***
	(16.79)	(12.00)	(9.64)	(10.77)
-SD	1.61***	1.51***	1.42***	1.51***
	(15.34)	(10.79)	(9.04)	(10.06)
p(sd)	-0.01	-0.24*	-0.65***	-0.32**
	(-0.08)	(-2.07)	(-4.81)	(-2.66)
C	-0.47***	-0.61***	-0.52***	-0.65***
	(-5.83)	(-5.67)	(-4.25)	(-5.65)
p(ep)	-1.17***	-0.95***	-0.36*	-1.02***
	(-12.02)	(-7.58)	(-2.51)	(-7.64)
residuals (performance model)		3.10***		
		(27.42)		
residuals (motivation model)			3.77***	
			(28.08)	
residuals (full model)				3.06***
				(25.72)
Nagelkerke's R^2	0.31	0.60	0.67	0.59
N	3206	2866	2621	2512

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.

(b) Synthetic file.

Predicted probabilities Following the proposition made by Lucas (2001, 2009), for the analysis of transition rate changes, predicted probabilities rather than regression coefficients should be compared. Figure 8 plots the predicted probabilities for students' educational transitions conditional on their teachers' residualized expectations (with the respective SEU indicators held constant at their mean). The solid lines show students' predicted transition probabilities conditional on their primary school teachers' transition recommendations, the dashed lines show students' predicted transition probabilities conditional on their secondary school teachers' evaluations. Apparently, differences in transition probabilities between students with primary school over- and underestimations, respectively, become smaller from transition 1 (+/- Gymnasium) to transition 2 (+/- Abitur). Likewise, differences between students with secondary school over- and underestimations, respectively, become smaller from transition 2 to transition 3 (+/- university transition). Both results provide support for hypothesis H_{2a}.

What does not suit this general trend is the finding that differences between students who had been over- and underestimated by their primary school teachers, respectively, seem to become larger again from transition 2 to transition 3; and that students who had been *under*estimated by their primary school teachers have a *higher* predicted university transition probability on average than those who had been *over*estimated by their primary school teachers. Yet, we have to keep in mind that the underlying logit coefficients are not significant – so this would still be weak evidence to reject both H_1 and H_{2b} in this respect.

On the other hand, the predicted university transition probabilities of students who had been overestimated at secondary school are still higher than the predicted probabilities of students who had been overestimated at primary school. This is well in line with hypothesis H_{2b}.

4.3 Robustness Analysis

In order to get an approximate intuition about the reliability of the results gained from statistical matching, I intend to perform a couple of robustness analyses. The latter cover i) a test for unobserved heterogeneity in terms of a potentially confounding variable; ii) additional secondary-school controls for the analysis of long-term primary school expectancy effects; and iii) a comparison of the empirical patterns in the synthetic file with results from a 'real' panel study with all required indicators in the same file (though less perfectly measured).

Are results robust against unobserved heterogeneity? First, I want to keep a check on the robustness of the results against an unobserved but potentially confounding variable u. Following Buis (2010, 2011), this kind of unobserved heterogeneity can be simulated by introducing a weighted sum of random variables $\nu_k = \beta_{uk}u$ – while k denotes several educational transition points – which is approximated by a normal distribution. The researcher can now simulate different values for both the standard deviation $sd(\nu)$ of this random variable and its correlation ρ with the predictor of interest. I used the `seqlogit` package (Buis, 2007) in Stata to simulate a scenario of $sd(\nu) = 0$ (the case

Table 5: Sequential logit model of students' educational transitions (synthetic data, *CHiSP* predictors)

	Model 5a	*Model 6a*	*Model 7a*	*Model 8a*	*Model 9a*	*Model 10a*	*Model 11a*	*Model 12a*
	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z
B	0.26**	0.10	0.12	0.23	-0.07	-0.48	-0.46	-0.46
	(2.71)	(0.56)	(0.68)	(1.47)	(-0.36)	(-1.37)	(-1.30)	(-1.31)
-SD	0.12	0.16	0.20	0.26	-0.18	-0.21	-0.18	-0.20
	(0.72)	(0.58)	(0.70)	(0.99)	(-0.58)	(-0.45)	(-0.40)	(-0.44)
p(sd)	-0.33	-0.30	-0.31	-0.40	-0.15	0.47	0.45	0.44
	(-1.83)	(-0.99)	(-1.02)	(-1.42)	(-0.42)	(0.78)	(0.75)	(0.73)
C	-0.11	0.22	0.22	0.21	0.10	0.03	0.02	0.04
	(-0.67)	(0.77)	(0.75)	(0.76)	(0.29)	(0.05)	(0.05)	(0.09)
p(ep)	1.60***	1.17	1.19	1.72**	-0.04	1.06	1.12	1.24
	(4.21)	(1.88)	(1.92)	(2.90)	(-0.04)	(0.79)	(0.84)	(0.99)
residuals (performance model)		1.83***				0.14		
		(6.28)				(0.21)		
residuals (motivation model)			1.78***				0.06	
			(6.17)				(0.09)	
residuals (full model)				0.76**				-0.11
				(2.91)				(-0.22)
Nagelkerke's R^2	0.06	0.25	0.25	0.14	0.06	0.25	0.25	0.14
N	725	297	297	294	725	297	297	294

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5a-8a = Abitur; models 9a-12a = university transitions.

Table 6: Sequential logit model of students' educational transitions (synthetic data, *PHEO* predictors)

	Model 5b	*Model 6b*	*Model 7b*	*Model 8b*	*Model 9b*	*Model 10b*	*Model 11b*	*Model 12b*
	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z
B	0.13	0.16	0.11	0.15	0.07	-0.09	-0.10	-0.11
	(0.98)	(1.18)	(0.73)	(0.96)	(0.26)	(-0.29)	(-0.31)	(-0.35)
-SD	-0.13	-0.17	-0.18	-0.22	0.51	0.53	0.48	0.46
	(-0.98)	(-1.17)	(-1.14)	(-1.33)	(1.54)	(1.44)	(1.28)	(1.19)
p(sd)	0.13	0.20	0.20	0.23	0.04	0.30	0.19	0.25
	(0.89)	(1.29)	(1.26)	(1.38)	(0.14)	(0.88)	(0.57)	(0.71)
C	0.25	0.31*	0.34*	0.36*	0.10	0.28	0.20	0.21
	(1.85)	(2.17)	(2.30)	(2.38)	(0.33)	(0.88)	(0.62)	(0.65)
p(ep)	0.04	-0.05	-0.10	-0.07	0.29	0.50	0.05	0.37
	(0.22)	(-0.25)	(-0.47)	(-0.34)	(0.66)	(1.00)	(0.11)	(0.72)
residuals (performance model)		-0.28				-0.51		
		(-1.28)				(-0.92)		
residuals (motivation model)			-0.35				-0.81	
			(-1.64)				(-1.42)	
residuals (full model)				-0.34				-0.58
				(-1.54)				(-1.04)
Nagelkerke's R^2	0.01	0.02	0.02	0.03	0.01	0.02	0.02	0.03
N	945	846	767	738	945	846	767	738

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5b-8b = Abitur; models 9b-12b = university transitions.

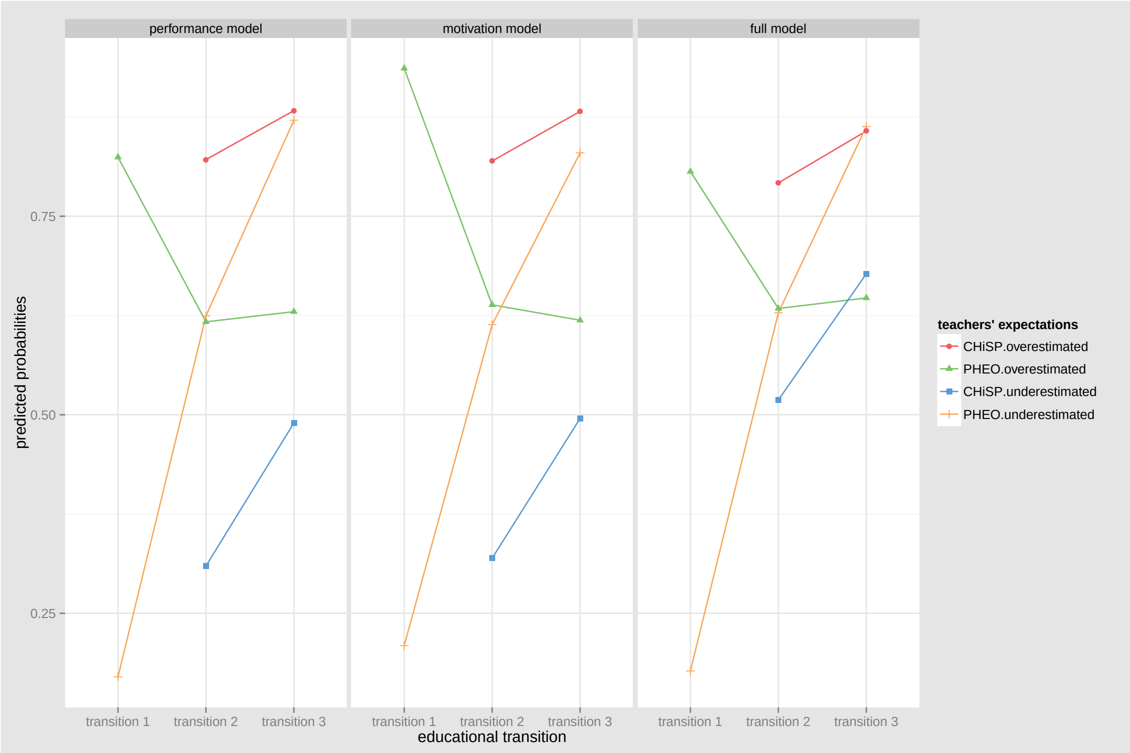


Figure 8: Predicted probabilities of students' educational transitions conditional on teachers' over- and underestimations.

of no unobserved covariate), $sd(\nu) = 0.5$, $sd(\nu) = 1$, and $sd(\nu) = 2$. In addition, ρ is allowed to vary between 0 up to .5. Figures 9 and 10 plot the estimated logit coefficients for the self-fulfilling prophecy residuals over educational transitions.

With regard to primary school teachers' expectations (fig. 9), the results indicate that a normally-distributed unobserved random variable with standard deviations up to 2 and correlations with the self-fulfilling prophecy residuals up to .5 would not be able to affect the confidence intervals (and thus the significance value) of the self-fulfilling prophecy residuals for the first transition. However, in the case of the second and the third transition, an insignificant coefficient could become significant if an unobserved variable would be correlated strongly enough with both the self-fulfilling prophecy residuals and the outcome of students' educational transitions: If the intercorrelations between the unobserved random variable and the self-fulfilling prophecy residuals are low, a standard deviation of $sd(\nu) = 2$ might cause an insignificant estimate to become positively significant. If the aforementioned intercorrelations are high (e.g. $\rho = .5$), even lower standard deviations would suffice an insignificant coefficient to become negatively significant.

Concerning secondary school teachers' expectations (fig. 10), in the case of high school graduations, it might happen that the significant estimate could become insignificant if the intercorrelations between the self-fulfilling prophecy residuals and the unobserved random variable are at least moderate ($\rho > .3$), and at the same time, the standard deviation of the latter is high ($sd(\nu) = 2$). Contrarily, in the case of university tran-

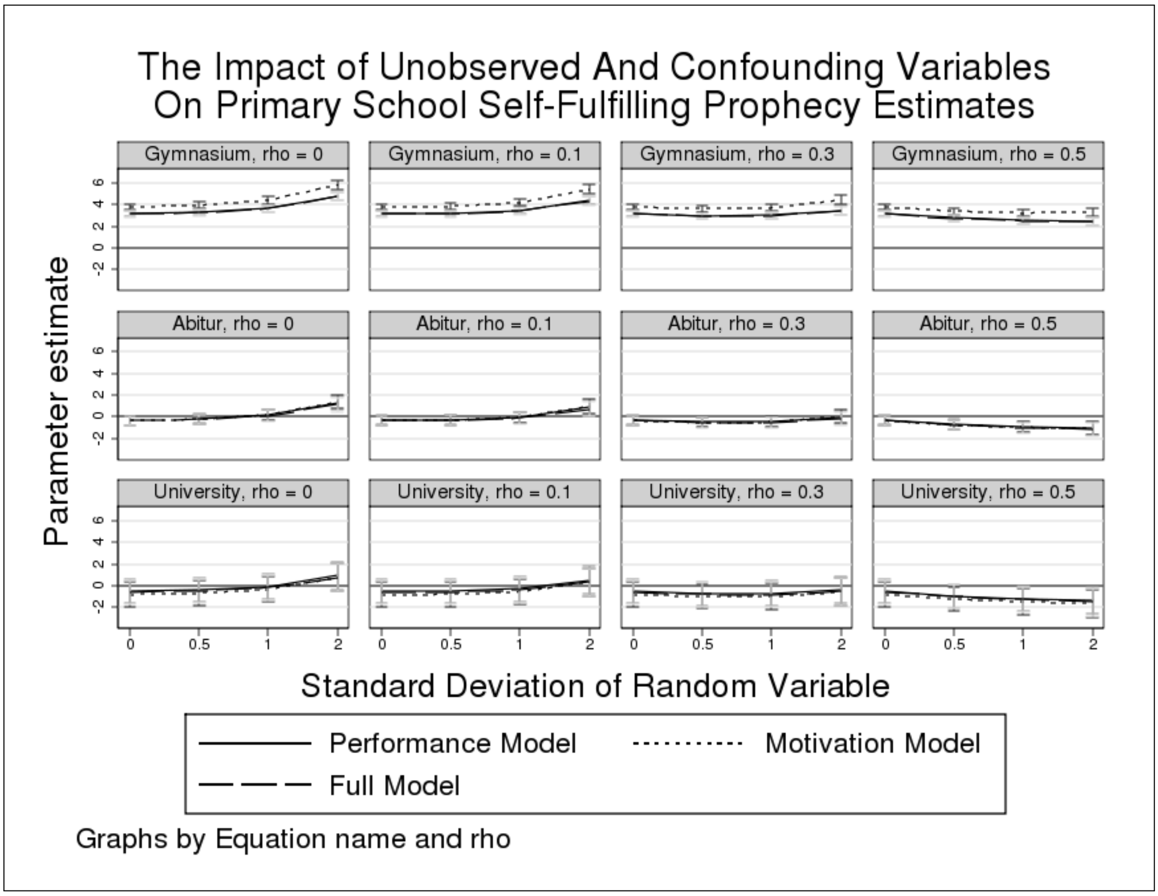


Figure 9: Sensitivity analysis, *PHEO* indicators.

sitions, it might happen that the insignificant self-fulfilling prophecy residuals could become significant again: Given a random variable with a comparably high standard deviation ($sd(\nu) = 2$), an insignificant estimate could become positively significant if its correlation ρ with ν is zero, and negatively significant if this correlation is high ($\rho = .5$).

Taken together, if the time interval between a teacher's expectation and a student's educational transition is short (such as in the case of primary school teachers' transition recommendations), a confounding variable is unlikely to affect its significant effect. The later an educational transition, and the larger the time span between a teacher's expectation and a student's transition, the more unobserved heterogeneity might be a problem.

Are results robust against additional secondary-school controls? One strategy to deal with the problem of unobserved heterogeneity is to introduce additional controls in the sequential logit equations. Regarding the long-term effects of primary school teachers' expectations, both secondary-school SEU terms and secondary-school expectancy effects might be indicators potentially associated with either the predictors of interest, the outcome, or both of them. Therefore, in figure 11, I present four different plots of

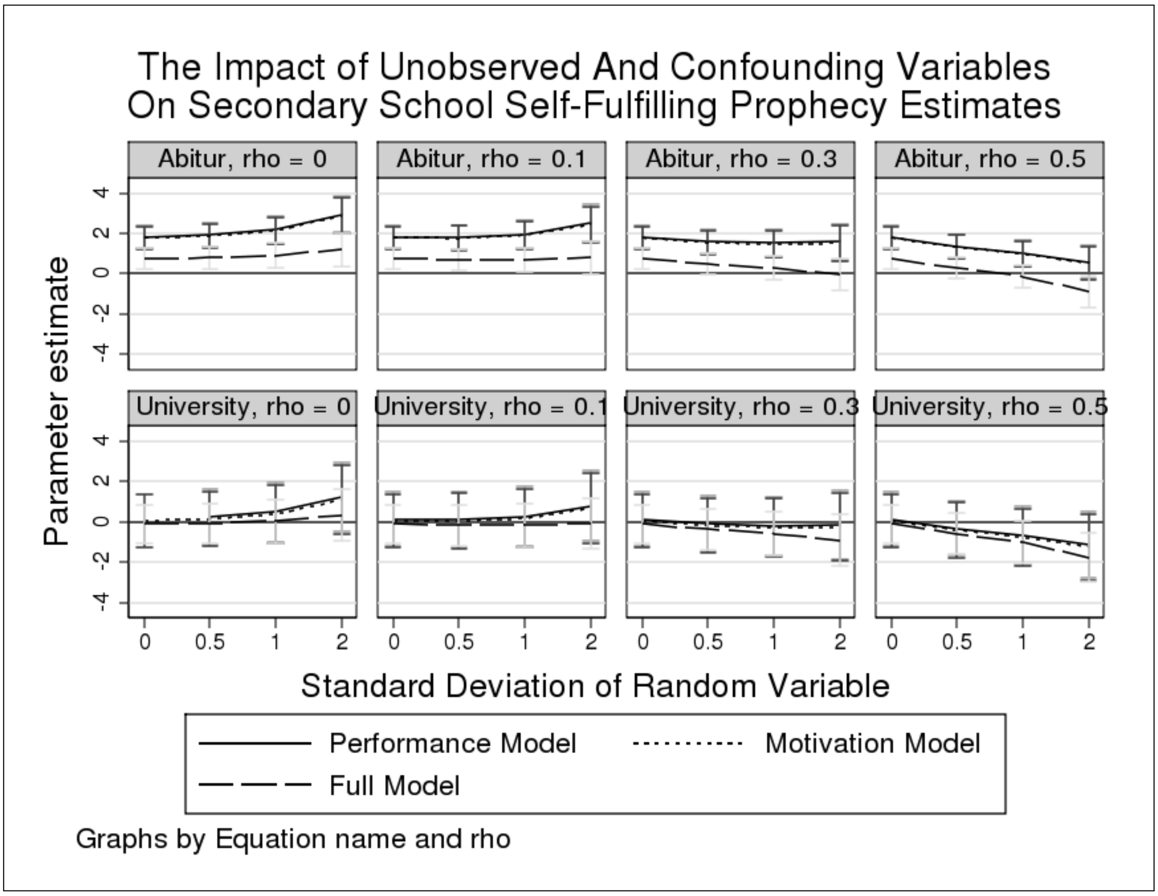


Figure 10: Sensitivity analysis, *CHiSP* indicators.

predicted probabilities of students' educational transitions depending on whether they had been over- or underestimated by their primary school teacher: The predicted probabilities in figure 11a do not depend on additional controls apart from the variables from table 6, figure 11b adds controls for secondary-school SEU terms, 11c adds controls for secondary-school teacher expectancy effects, and figure 11d adds controls for both secondary school SEU terms and teacher expectancy effects.[38]

[38]The underlying logit models of figures 11b – 11d are shown in tables H – J.

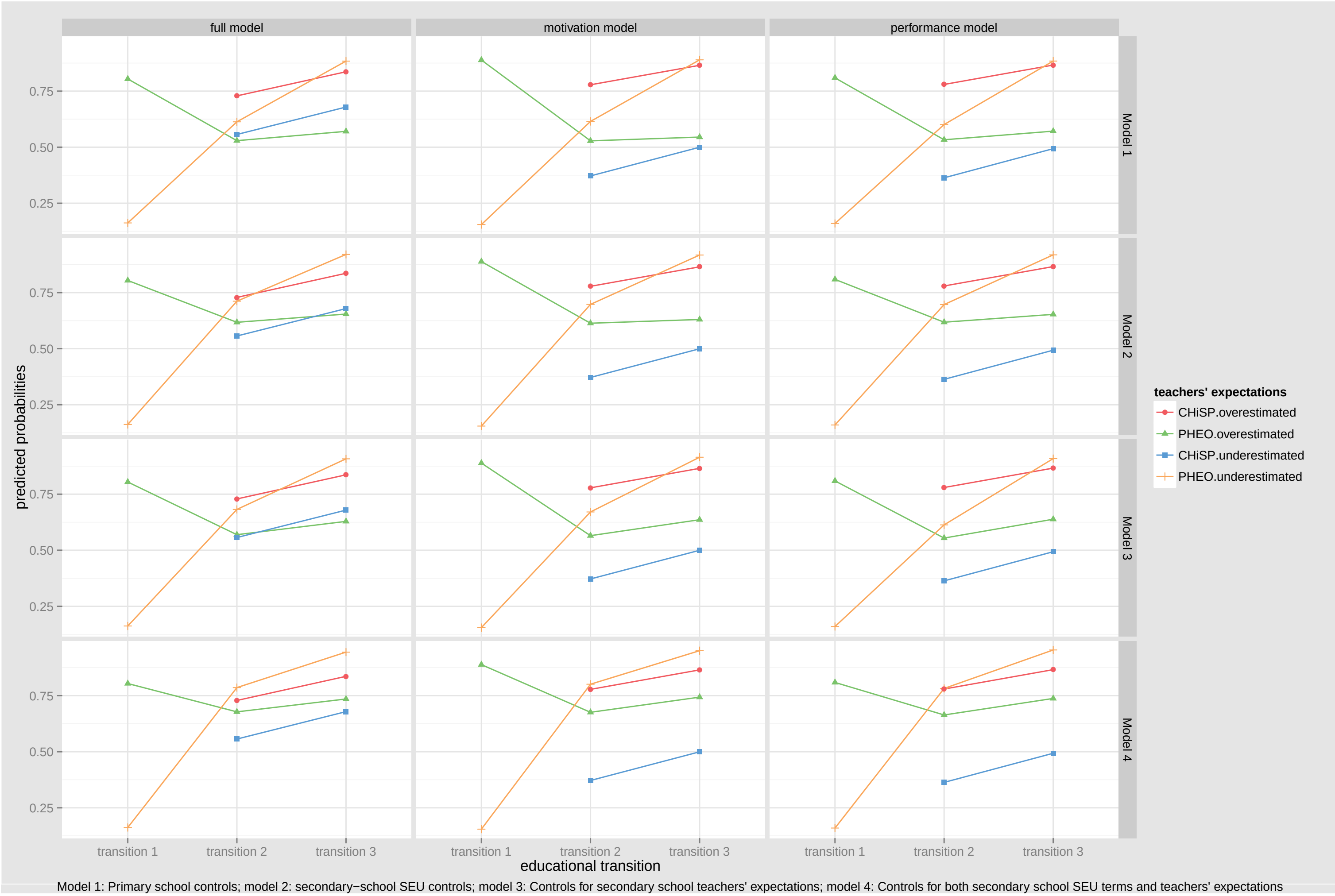


Figure 11: A robustness analysis of predicted probabilities of educational transitions.

We can observe that introducing the above-mentioned controls only marginally affects the predicted transition probabilities for over- and underestimated students, respectively. A pleasing result is that the counter-intuitively positive (though insignificant) difference in predicted probabilities between under- and overestimated students at t_2 and t_3 becomes smaller with controls for both secondary-school SEU terms and residualized teacher expectations. On the other hand, also the hypothesized positive difference between later and earlier overestimations at a given transition point is reduced – though not as strongly as that H_{2b} would have to be rejected.[39]

Are results from statistical matching comparable to 'real' panel analysis? Finally, I compare the results from the synthetic file with analyses based on an actual, non-matched student cohort. For this purpose, I draw on the *B*ritish *C*ohort *S*tudy (*BCS*) that follows a birth cohort of children all born in the same week in April 1970. For the robustness analyses to be conducted here, I formed SEU indicators from both student- and parent-level re-surveys from 1980 and 1986. (As mentioned, I consider these indicators to be less precise than in the synthetic file – which is why I still believe statistical matching of *PHEO* and *CHiSP* to yield more convincing results.) Most importantly, in the above-mentioned time span, students' current teachers were surveyed as well, and they were asked various evaluative questions from which I could derive tolerably comparable teacher expectancy indicators. Moreover, from the (former) student re-survey conducted in 2000, I could add information about students' educational transitions.[40] A summary of the British cohort data variables used for the analyses at hand as well as their distribution is given in tables K, L, and M (appendix).

Figure 12 graphs the predicted educational transition probabilities for those students who had been over- and underestimated at age 10 and 16, respectively. Unlike in the synthetic *PHEO-CHiSP* file, now we also observe the effect of secondary-school teachers' expectations on the first transition since the first possible transition to observe (O level vs. no qualification) lies temporarily behind the survey time of the second expectancy indicator.[41]

For the *BCS* data, the picture is a bit more puzzling than for the synthetic file. First, already at t_1, differences in predicted transition probabilities between over- and underestimated students are notably smaller than in the graphs based on the synthetic file, and they appear to diminish entirely at t_3. Second, differences between students who had been overestimated by secondary school teachers and students who had been overestimated by primary school teachers are also smaller and tend to disappear at

[39] Looking at the underlying logit models in table J (appendix), we can observe that when both primary and secondary school teachers' expectations are in the model, two out of three indicators still significantly predict students' propensity to pass Abitur.

[40] Note that the British educational system differs remarkably from the German educational system since in the former, students are not tracked after primary school. Therefore, the dependent variable of the analyses at hand changes to the following transitions: 0 'no qualification'; 1 '+/- O-level GCE'; 2 '+/- A-level GCE' and 3 '+/- university transition'. This, of course, might harden the comparability between the two data sets. Furthermore, in the *BCS*, it is not possible to observe when students started academic studies. Thus, I was not able to set a cut-off like in *CHiSP*.

[41] The underlying logit models are shown in tables N and O (appendix).

t_3. Although on the one hand, this could be in line with the dissipation hypothesis in self-fulfilling prophecy research (due to the lacking possibility to specify a cutoff as in *CHiSP*), on the other hand, we cannot exclude that this result is due to institutional or measurement differences between the *BCS* and the (German) synthetic file. But since apart from the data points for t_3, the general pattern of figure 12 is in line with my expectations and the underlying self-fulfilling prophecy estimates lack statistical significance, one can conclude that the results from the synthetic file are sufficiently robust.

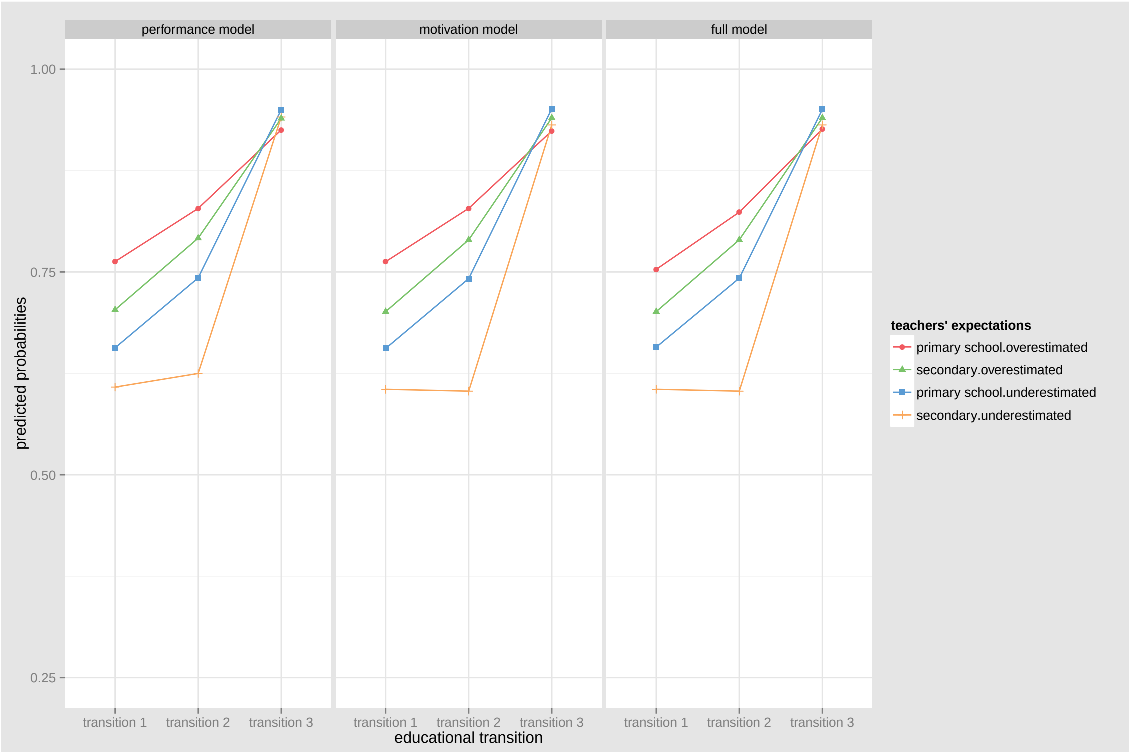


Figure 12: Predicted probabilities of students' educational transitions conditional on teachers' over- and underestimations (British Cohort Study).

5 Conclusion

Regarding theoretical advancement, the paper at hand is intended to unify findings from rational-choice oriented educational sociology and *Pygmalion* as well as self-fulfilling prophecy research (Rosenthal and Jacobson, 1968; Raudenbush, 1984; Jussim, 1986; Jussim and Harber, 2005) about declining effect sizes over students' educational transitions. While conventional rational choice explanations such as the *L*ife *C*ourse *P*erspective (*LCP*; Müller and Karle, 1993), *M*aximally *M*aintained *I*nequality (*MMI*; Raftery and Hout, 1993), *R*elative *R*isk *A*version (*RRA*; Breen and Goldthorpe, 1997) or *E*ffectively *M*aintained *I*nequality (*EMI*; Lucas, 2001) usually remain restricted to student- or parent-level utility considerations, a main advantage of social-psychological theories, though less rich with respect to the former indicators, is their explicit consideration of teacher expectancy effects. In the theoretical section, I first alluded to the main contributions from both disciplines to explain decreasing background effects over students' educational life course.

Starting from a recent *s*ubjective *e*xpected *u*tility (SEU) explanation of how teachers' expectations of their students can be supposed to affect the latter's transition propensities via their subjective expected probabilities of educational success (Becker, 2010a; also see Esser, 1999), I then argued that via a mechanism of *Bayesian updating* (Breen, 1999; Breen and García-Peñalosa, 2002; Morgan, 2005, ch. 5), students adjust their beliefs after both a teacher's evaluation and an educational transition. By means of a simplified decision tree, I endeavored to show how belief updating could explain the empirically-observed phenomenon of decreasing teacher expectancy effects over time – after which I deduced corresponding hypotheses suiting the conventional framework of educational transition analysis.

Regarding methodological advancement, teacher expectancy effects were then operationalized as over- and underestimations (in terms of residualized expectations) in a repeated belief model (i.e. with measures at two distinct time points). As it has been argued elsewhere (Becker, 2010a), these residuals can be understood as an approximation of the unobserved mechanisms that account for teacher treatment effects on students' subjective expected probability of educational success.

Since I did not find a single file that could have been used to test a repeated belief model with convincing controls for conventional SEU terms, I used the *distance hot deck method* (see D'Orazio et al., 2006) to create an artificial student cohort by means of statistical matching (Rubin, 1986). Concretely, one source which measured teachers' transition recommendations at the end of primary school (Gesis-No. ZA893) was matched to a second source that surveyed teachers' secondary school evaluations of students' prospective academic ability (Gesis-No. ZA640). Before this was performed, I justified how my theoretical model can be supposed to fulfill the important *c*onditional *i*ndependence *a*ssumption (CIA).

From several sequential logit models we could note that students who were overestimated by their teachers in general have a significantly higher propensity of taking the next educational transition. This is well in line with conventional self-fulfilling prophecy and *Pygmalion* research (for research overviews see Raudenbush, 1984; Jussim, 1986; Rosenthal, 1994; Jussim and Harber, 2005) and its recent transformations on more sociological questions (Becker, 2010a). Regarding *changes* in this effect over time, I found evidence that i) differences in predicted transition probabilities between students with primary school over- and underestimations, respectively, become smaller from transition 1 (+/- Gymnasium) to transition 2 (+/- Abitur); ii) differences in predicted transition probabilities between students with secondary school over- and underestimations, respectively, become smaller from transition 2 to transition 3 (+/- university transition); and iii) the predicted transition probabilities of students who had been overestimated at secondary school to make the transition to university are still higher than the predicted transition probabilities of students who had been overestimated at primary school. Regarding the implied mechanism of Bayesian updating that was postulated in section 2.4, it was illustrated that if differences in student beliefs after an over- or underestimation tended to decrease over their educational life course, one would be justified to conclude that long-term teacher expectancy effects on student beliefs are smaller than short-term effects. Thus, by trend, the results at hand support my hypotheses.

In several robustness analyses, I first noted that the later an educational transition, and the more time falls between a teacher's expectation and a student's educational transition, the more the residualized teacher expectancy effects might stand at risk of being affected by an unobserved but confounding variable. What goes in line with that observation is that the counter-intuitive (though insignificant) result that students who had been *under*estimated at primary school show *higher* predicted probabilities of university transitions than students who had been *over*estimated was in part canceled out with additional controls for both secondary school teachers' expectations and student SEU terms.

With respect to the assumption of Bayesian updating, this result could be a hint that the simplified decision tree that was proposed in section 2.4 only provides a rough sketch of the social situation; and apart from regarding only isolated over- and underestimations, the whole sequence of expectancy effects including students' cost-benefit evaluations at all available measurement points should be considered.

A final comparison with the British Cohort Study data (that I consider to entail indicators less perfectly measured) showed that although teacher expectancy effects were generally lower than in the synthetic file, the general pattern of declining effects over time also undergirds the reliability of the results from the synthetic file.

Overall, these findings suggests the following improvements: From a theoretical point of view, a comprehensive Bayesian analysis of the underlying educational decision tree would certainly be insightful, but would have gone far beyond the present study's scope. In my view, Breen (1999), Breen and García-Peñalosa (2002) and Morgan (2005, ch. 5) have initiated an important theoretical advancement in educational sociology that should be generalized to include all relevant parameters affecting students' utility function that may be subject to updating processes over their educational life course. Once this has been performed for teacher expectancy effects, the model could be fitted in a corresponding Bayesian probit model of choice making (Albert and Chib, 1993; Jackman, 2009, ch. 8).

Furthermore, as acknowledged in the operationalization section, all results have to be evaluated cautiously. Concerning the statistical matching procedure, admittedly, the two sources stem from different populations since in the file "Parental Home and Educational Opportunities" (*PHEO*), primary school students in Baden-Wuerttemberg were surveyed, while in the Cologne High School Panel (*CHiSP*), secondary school students in North Rhine-Westphalia were interviewed. Thus, institutional factors in the two German Federal Lands may lead to difficulties in comparability since, in technical terms, the two sources are *biased* (van der Laan, 2000). A potential strategy to cope with this problem could be to use another measure of primary school teachers' expectations such as general ability assessments which do not suffer from institutional dependencies. A second problem may arise due to measurement differences in the two data files. In particular, it was not possible to compute a high estimated benefit of education for parents in *PHEO* – which may be an issue of *inconsistency* as admonished by van der Laan (2000) –, but a comparison of logistic regression estimates based on both original *PHEO* and *PHEO* indicators in the synthetic file revealed differences in significance also for other coefficients. Thus, I would still recommend to treat the present results with necessary prudence – though the estimates obtained from the British Cohort study basically tend into the same direction.

A more technical recommendation relates to the choice of the concrete matching algorithm. In the study at hand, I chose an unconstrained *distance hot deck* algorithm relying on the Gower distance function for both binary and ordered categorical levels of measurement. However, future studies should test whether results are reproducible if alternative matching algorithms as reviewed in section 3.3 are applied.

In the context of further improvements, a special situation could occur in terms of a sequence of multiple underestimations: According to findings by Madon et al. (2006), it is be possible that self-fulfilling prophecies could accumulate rather than dissipate over time. Hence, in further analyses, it should also be tested if the predicted transition probabilities of students who had been underestimated by their teachers at different time points are lower than they would be had only a single teacher expectation been considered. Moreover, as some self-fulfilling prophecy studies suggest (Jussim and Harber, 2005), additional interactions could also be formed with students' social backgrounds. And finally, future studies should extend their analysis of changes in predicted probabilities of teachers' expectations over educational transitions *across cohorts*. Since most of the above-cited rational choice theories also address this issue, it would be an interesting but neglected question also for self-fulfilling prophecy research. However, answering this question would even more depend on data availability.

References

Albert JH and Chib S (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88(422): 669–679.

Becker D (2010a) The impact of teachers' expectations on students' educational opportunities in the life course. *Cologne Graduate School Working Paper Series* 1(7).
URL: *http://econpapers.repec.org/paper/cgrcgsser/01-07.htm*

Becker D and Birkelbach K (2011) Teachers' evaluations and the definition of the social situation in the classroom. *Cologne Graduate School Working Paper Series* 2(4).
URL: *http://ideas.repec.org/p/cgr/cgsser/02-04.html*

Becker R (2003) Educational expansion and persistent inequalities of education. *European Sociological Review* 19(1): 1–24.

Becker R (2010b) Soziale Ungleichheit von Bildungschancen und Chancengerechtigkeit - eine Reanalyse mit bildungspolitischen Implikationen. In: Becker R and Lauterbach W (eds.) *Bildung als Privileg*, 161–189. Wiesbaden: VS Verlag für Sozialwissenschaften.

Becker R and Hecken AE (2007) Studium oder Berufsausbildung? Eine empirische Überprüfung der Modelle zur Erklärung von Bildungsentscheidungen von Esser sowie von Breen und Goldthorpe. *Zeitschrift für Soziologie* 36(2): 100–117.

Becker R and Hecken AE (2009a) 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.

Becker R and Hecken AE (2009b) Why are working-class children diverted from universities? An empirical assessment of the diversion thesis. *European Sociological Review* 25(2): 233–250.

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

Breen R (1999) Beliefs, rational choice and bayesian learning. *Rationality and Society* 11(4): 463–479.

Breen R and García-Peñalosa C (2002) Bayesian learning and gender segregation. *Journal of Labor Economics* 20(4): 899–922.

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

Breiman L; Friedman JH; Olshlen RA et al. (1984) *Classification and Regression Trees.* Belmont, CA: Wadsworth.

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.

Buis ML (2011) The consequences of unobserved heterogeneity in a sequential logit model. *Research in Social Stratification and Mobility* 29(3): 247–262.

Caliendo M and Kopeinig S (2008) Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22(1): 31–72.

Cameron SV and Heckman JJ (1998) Life cycle schooling and dynamic selection bias: models and evidence for five cohorts of American males. *Journal of Political Economy* 106(2): 262–333.

Cobalti A (1990) Schooling inequalities in Italy: trends over time. *European Sociological Review* 6(3): 199–214.

de Boer H; Bosker RJ and van der Werf MPC (2010) Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology* 102(1): 168–179.

De'ath G and Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81(11): 3178–3192.

D'Orazio M (2009) *StatMatch: Statistical Matching. R package version 0.8.* **URL:** *http://cran.r-project.org/package=StatMatch*

D'Orazio M; Di Zio M and Scanu M (2006) *Statistical Matching - Theory and Practice.* Chichester: John Wiley & Sons.

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.

Erikson RS and Jonsson JO (1996) Introduction: explaining class inequality in education: the Swedish test case. In: Erikson RS and Jonsson JO (eds.) *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, 1–63. Boulder, CO: Westview Press.

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

Frieze IH; Olson JE and Russell J (1991) Attractiveness and income for men and women in management. *Journal of Applied Social Psychology* 21(13): 1039–1057.

Ganzeboom HBG and Treiman DJ (1996) Internationally comparable measures of occupational status for the 1988 International Standard Classification of Occupations. *Social Science Research* 25(3): 201–239.

Gill S and Reynolds AJ (1999) Educational expectations and school achievement of urban African American children. *Journal of School Psychology* 37(4): 403–424.

Hansen M; Dubayah R and DeFries R (1996) Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing* 17(5): 1075–1081.

Hinnant JB; O'Brien M and Ghazarian SR (2009) The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology* 101(3): 662–670.

Ingram DD; O'Hare J; Scheuren F et al. (2000) Statistical matching: a new validation case study. *Proceedings of the Survey Research Methods Section, American Statistical Association* 746–751.

Jackman S (2009) *Bayesian Analysis for the Social Sciences.* Chichester: Wiley.

Jussim L (1986) Self-fulfilling prophecies: a theoretical and integrative review. *Psychological Review* 93(4): 429–445.

Jussim L (1989) Teacher expectations: self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology* 57(3): 469–480.

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.

Little RJA and Rubin DB (2002) Statistical Analysis With Missing Data. *Hoboken (NJ): Wiley-Interscience* .

Lucas SR (2001) Effectively maintained Iinequality: education transitions, track mobility, and social background effects. *American Journal of Sociology* 106(6): 1642–1690.

Lucas SR (2009) Stratification theory, socioeconomic background, and educational attainment. *Rationality and Society* 21(4): 459–511.

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; Willard J; Guyll M et al. (2006) Self-fulfilling prophecy effects of mothers' beliefs on children's alcohol use: accumulation, dissipation, and stability over time. *Journal of personality and social psychology* 90(6): 911–926.

Manski CF (2004) Measuring expectations. *Econometrica: Journal of the Econometric Society* 72(5): 1329–1376.

Mare RD (1980) Social background and school continuation decisions. *Journal of the American Statistical Association* 75(370): 295– 305.

Mare RD (1981) Change and stability in educational stratification. *American Sociological Review* 46(1): 72–87.

Mechtenberg L (2009) Cheap talk in the classroom: How biased grading at school explains gender differences in achievements, career choices and wages. *Review of Economic Studies* 76(4): 1431–1459.

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

Mistry RS; White ES; Benner AD et al. (2009) A longitudinal study of the simultaneous influence of mothers' and teachers' educational expectations on low-income youth's academic achievement. *Journal of Youth and Adolescence* 38(6): 826–838.

Morgan SL (2002) Modeling preparatory commitment and non-repeatable decisions. *Rationality and Society* 14(4): 387–429.

Morgan SL (2005) *On the Edge of Commitment.* Stanford: Stanford University Press.

Muller C; Katz SR and Dance LJ (1999) Investing in teaching and learning: dynamics of the teacher-student relationship from each actor's perspective. *Urban Education* 34(3): 292–337.

Müller W and Karle W (1993) Social selection in educational systems in Europe. *European Sociological Review* 9(1): 1–23.

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

Nash R (2003) Inequality/difference in education: is a real explanation of primary and secondary effects possible? *The British Journal of Sociology* 54(4): 433–451.

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.

Prasad AM; Iverson LR and Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9(2): 181–199.

Raftery AE and Hout M (1993) Maximally maintained inequality: expansion, reform, and opportunity in Irish education, 1921-75. *Sociology of Education* 66(1): 41–62.

Rässler S (2002) *Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches.* Berlin: Springer.

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.

Ripley B (2010) *tree: Classification and regression trees. R package version 1.0-28.* Tech. rep.
URL: *http://cran.r-project.org/package=tree*

Rist RC (1970) Student social class and teacher expectations: the self-fulfilling prophecy in ghetto education. *Harvard Educational Review* 40(3): 411–451.

Rosenthal R (1994) Interpersonal expectancy effects: a 30-year perspective. *Current Directions in Psychological Science* 3(6): 176–179.

Rosenthal R and DiMatteo MR (2001) Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology* 52: 59–82.

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.

Rubin DB (1986) Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business & Economic Statistics* 4(1): 87–94.

Schubert F and Becker R (2010) Social inequality of reading literacy - a longitudinal analysis with cross-sectional data of PIRLS 2001 and PISA 2000 utilizing the pair wise matching procedure. *Research in Social Stratification and Mobility* 28(1): 109–133.

Sewell WH; Haller AO and Ohlendorf GW (1970) The educational and early occupational status attainment process: replication and revision. *American Sociological Review* 35(6): 1014–1027.

Sewell WH; Haller AO and Portes A (1969) The educational and early occupational attainment process. *American Sociological Review* 34(1): 82–92.

Siddique J and Belin TR (2008) Multiple imputation using an iterative hot-deck with distance-based donor selection. *Statistics in Medicine* 27(1): 83–102.

Smith AE; Jussim L and Eccles J (1999) Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time? *Journal of Personality and Social Psychology* 77(3): 548–565.

Smith HL and Cheung PPL (1986) Trends in the effects of family background on educational attainment in the Philippines. *American Journal of Sociology* 91(6): 1387–1408.

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

Tolsma J; Coenders M and Lubbers M (2007) Trends in ethnic educational inequalities in the Netherlands: a cohort design. *European Sociological Review* 23(3): 325–339.

van der Laan P (2000) Integrating administrative registers and household surveys. *Netherlands Official Statistics* 15(special issue): 7–15.

West CK and Anderson TH (1976) The question of preponderant causation in teacher expectancy research. *Review of Educational Research* 46(4): 613–630.

6 Appendix

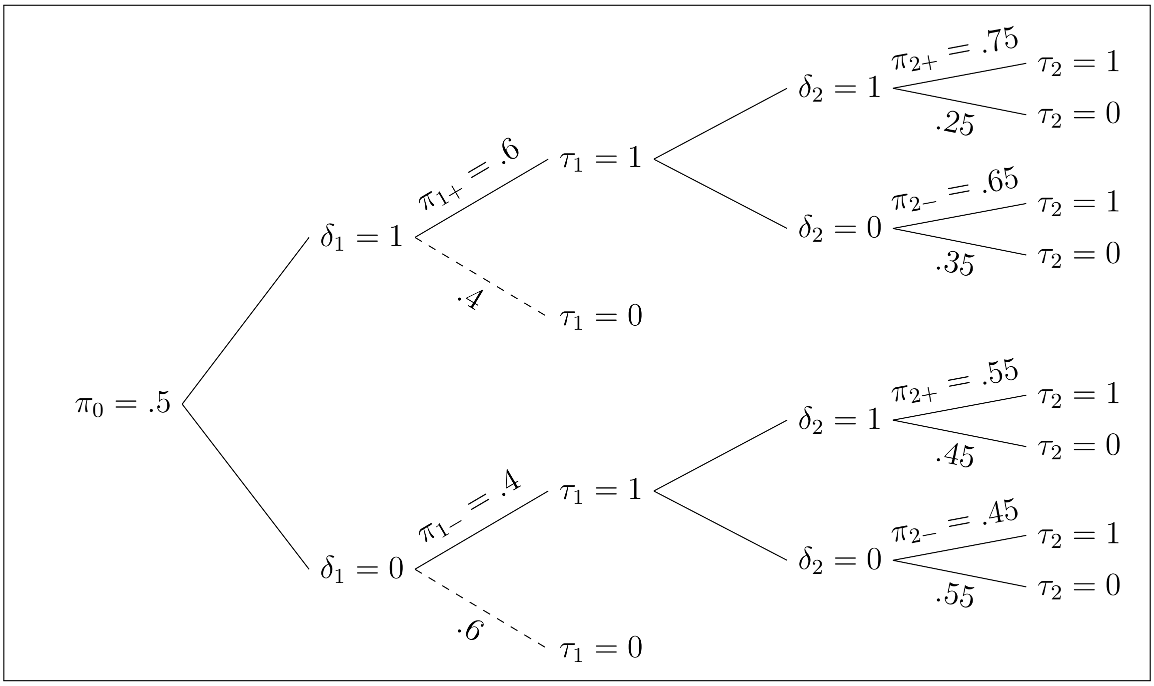


Figure A: Decision tree under scenario 3 (see table 1).

Table A: Logistic regression model of primary school teachers' recommendations on students' performance and motivation (original *PHEO* data)

	Performance Model e^{b*sd}/z-value	*Motivation Model* e^{b*sd}/z-value	*Full Model* e^{b*sd}/z-value
average grade	3.80***		3.76***
	(15.71)		(14.41)
learning motivation		1.37***	0.98
		(5.54)	(-0.37)
TV consumption		1.10	1.05
		(1.80)	(0.77)
Nagelkerke's R^2	0.29	0.03	0.29
N	1549	1453	1393

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

Table B: Logistic regression model of secondary school teachers' evaluations on students' performance and motivation (original *CHiSP* data)

	Performance Model e^{b*sd}/z-value	*Motivation Model* e^{b*sd}/z-value	*Full Model* e^{b*sd}/z-value
average grade	8.40***		7.45***
	(17.63)		(16.41)
learning motivation		1.14*	1.11
		(2.21)	(1.37)
self-concept		2.03***	1.45***
		(10.67)	(4.52)
Nagelkerke's R^2	0.51	0.15	0.52
N	1313	1304	1301

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

Table C: Classification of occupations in *PHEO* according to the prestige score provided by Ganzeboom and Treiman (1996)

	leitende und wiss. Angestellte	mittl. Angestellte	untergeordnete Angestellte	Beamte des höheren Dienstes	Beamte des gehobenen und mittleren Dienstes	Beamte des einfachen Dienstes	freie Berufe	Inhaber und Geschäftsführer von Großbetrieben	mittlere und kleine Selbständige	Selbständige Landwirte mit Großbetrieb	Selbstständige Landwirte mit mittelgroßem Betrieb	Facharbeiter mit abgelegter Prüfung	angelernte Arbeiter	Ungelernte und Hilfsarbeiter
Landwirtschaftliche Lehre	60				48				47	60	47	38	21	18
Industrielle und handwerkliche Lehre	60	52	38		48	37	52	60	52		47	38	21	18
Technische Berufe, Ingenieure, techn. Zeichner	63	56		48	72	56	36	63	56			38	21	
Berufe des Gaststättenwesens und priv. Dienstl.	60	37	26		49	37	44	60	44			38	21	18
Handels- und Verkehrsberufe	60	43	39				52		52			38	21	18
Berufe der Verwaltung, des Rechtswesens u. d. Sozialpflege	63	49	37	70	49	37	52	63	52			38	21	18
Berufe des Gesundheitswesens, Lehr- u. Erziehungsberufe	60	57	50	72	57	50	52		52			38	21	18
Arbeitskräfte mit unbestimmtem Beruf		49							52					
Anderes	na	na	na	na	na	na	na	na	na	na	na	na	na	na

Notes: Empty cells correspond to combinations of *status* and *situs* that were not observed in *PHEO*. Respondents who did not state an explicit occupation were set to missing. For the full list of occupational classification and the corresponding prestige score see Ganzeboom and Treiman (1996, pp. 221-237).

Table D: Summary statistics of all independent variables in the synthetic file

	count	mean	sd	min	max
B [PHEO]	3374	1.13	0.52	0.00	2.00
-SD [PHEO]	3358	0.24	0.43	0.00	1.00
p(sd) [PHEO]	3244	0.32	0.46	0.00	1.00
C [PHEO]	3339	0.48	0.50	0.00	1.00
p(ep) [PHEO]	3367	0.26	0.44	0.00	1.00
transition recommendation [PHEO]	3133	0.49	0.50	0.00	1.00
residual (perf.) [PHEO]	3001	0.09	0.43	-0.86	1.12
residual (mot.) [PHEO]	2748	0.10	0.49	-0.48	0.83
residual (full) [PHEO]	2634	0.07	0.43	-0.86	1.13
residual (perf, dichot.) [PHEO]	3001	0.56	0.50	0.00	1.00
residual (mot., dichot.) [PHEO]	2748	0.47	0.50	0.00	1.00
residual (full, dichot.) [PHEO]	2634	0.54	0.50	0.00	1.00
par. occ. pres. [PHEO]	3213	49.34	14.12	18.00	72.00
par. occ. pres. (dichot.) [PHEO]	3374	0.70	0.46	0.00	1.00
par. educ. [PHEO]	3374	1.14	1.23	0.00	3.00
-SD [CHiSP]	2349	0.41	0.49	0.00	1.00
C [CHiSP]	2688	0.46	0.50	0.00	1.00
p(sd) [CHiSP]	2667	0.37	0.48	0.00	1.00
p(ep) [CHiSP]	2688	0.95	0.23	0.00	1.00
B [CHiSP]	3214	1.13	0.86	0.00	2.00
residual (perf.) [CHiSP]	1307	0.01	0.38	-0.99	0.99
residual (mot.) [CHiSP]	1298	0.00	0.47	-0.85	0.94
residual (full) [CHiSP]	1281	-0.00	0.38	-1.20	1.05
residual (perf, dichot.) [CHiSP]	1307	0.55	0.50	0.00	1.00
residual (mot., dichot.) [CHiSP]	1298	0.55	0.50	0.00	1.00
residual (full, dichot.) [CHiSP]	1281	0.48	0.50	0.00	1.00
par. educ. [CHiSP]	3374	1.14	1.23	0.00	3.00
par. occ. pres. [CHiSP]	2680	49.39	12.64	18.00	78.00
par. occ. pres. (dichot.) [CHiSP]	2680	0.52	0.50	0.00	1.00

Table E: Logit model of secondary school transitions with controls for parental occupational prestige and education

	Model 1c log-odds/z	*Model 2c* log-odds/z	*Model 3c* log-odds/z	*Model 4c* log-odds/z
B	2.12***	2.16***	1.88***	2.19***
	(9.00)	(7.33)	(5.40)	(7.04)
-SD	1.40***	1.56***	1.42***	1.55***
	(6.45)	(5.68)	(4.34)	(5.33)
p(sd)	0.09	0.22	-0.08	0.09
	(0.47)	(0.96)	(-0.27)	(0.36)
C	-0.47**	-0.53*	-0.46	-0.51*
	(-2.75)	(-2.45)	(-1.66)	(-2.22)
p(ep)	-1.72***	-1.58***	-1.05***	-1.65***
	(-8.96)	(-6.70)	(-3.67)	(-6.65)
par. occ. pres.	0.02*	0.01	-0.00	0.01
	(2.25)	(0.74)	(-0.05)	(0.57)
par. educ.	0.64***	0.74***	0.79***	0.66***
	(4.64)	(4.08)	(3.92)	(3.50)
residuals (performance model)		2.74***		
		(11.97)		
residuals (motivation model)			4.10***	
			(13.67)	
residuals (full model)				2.67***
				(11.13)
Nagelkerke's R^2	0.45	0.63	0.76	0.63
N	876	783	727	694

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.

(a) Original *PHEO* data.

	Model 1d log-odds/z	*Model 2d* log-odds/z	*Model 3d* log-odds/z	*Model 4d* log-odds/z
B	1.07***	0.92***	0.75***	0.87***
	(10.47)	(7.33)	(5.23)	(6.68)
-SD	1.60***	1.46***	1.49***	1.47***
	(13.35)	(9.63)	(8.58)	(9.39)
p(sd)	0.05	-0.20	-0.58***	-0.28*
	(0.49)	(-1.54)	(-3.89)	(-2.17)
C	-0.34***	-0.46***	-0.38**	-0.48***
	(-3.63)	(-3.84)	(-2.79)	(-3.89)
p(ep)	-1.29***	-1.09***	-0.57***	-1.06***
	(-11.46)	(-7.91)	(-3.68)	(-7.43)
par. occ. pres.	0.04***	0.03***	0.03***	0.02***
	(10.33)	(5.68)	(5.13)	(4.87)
par. educ.	0.74***	0.77***	0.77***	0.70***
	(16.77)	(12.63)	(11.72)	(11.30)
residuals (performance model)		2.90***		
		(23.44)		
residuals (motivation model)			3.76***	
			(25.19)	
residuals (full model)				2.87***
				(22.45)
Nagelkerke's R^2	0.49	0.66	0.73	0.65
N	3050	2721	2511	2405

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.

(b) Synthetic file.

Table F: Sequential logit model of students' educational transitions with controls for parental occupational prestige and education (synthetic data, *CHiSP* predictors)

	Model 5c	*Model 6c*	*Model 7c*	*Model 8c*	*Model 9c*	*Model 10c*	*Model 11c*	*Model 12c*
	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z
B	0.21*	0.08	0.10	0.20	-0.06	-0.52	-0.50	-0.50
	(2.23)	(0.46)	(0.57)	(1.22)	(-0.30)	(-1.45)	(-1.40)	(-1.41)
-SD	0.09	0.15	0.18	0.23	-0.20	-0.21	-0.18	-0.20
	(0.54)	(0.51)	(0.63)	(0.85)	(-0.64)	(-0.44)	(-0.39)	(-0.44)
p(sd)	-0.22	-0.25	-0.25	-0.29	-0.18	0.60	0.59	0.58
	(-1.15)	(-0.77)	(-0.79)	(-0.94)	(-0.48)	(0.95)	(0.94)	(0.93)
C	0.03	0.28	0.27	0.33	0.07	0.17	0.17	0.19
	(0.14)	(0.89)	(0.88)	(1.12)	(0.21)	(0.33)	(0.33)	(0.37)
p(ep)	1.71***	1.23	1.25*	1.85**	-0.13	1.30	1.36	1.50
	(4.44)	(1.94)	(1.98)	(3.06)	(-0.12)	(0.93)	(0.98)	(1.15)
par. occ. pres.	0.01	0.00	0.00	0.01	-0.02	-0.00	-0.00	-0.00
	(1.34)	(0.06)	(0.08)	(0.63)	(-1.00)	(-0.11)	(-0.09)	(-0.07)
par. educ.	0.07	0.04	0.05	0.06	0.11	0.19	0.20	0.20
	(0.88)	(0.32)	(0.34)	(0.47)	(0.70)	(0.85)	(0.85)	(0.86)
residuals (performance model)		1.80***				0.12		
		(6.12)				(0.17)		
residuals (motivation model)			1.75***				0.04	
			(6.02)				(0.06)	
residuals (full model)				0.71**				-0.16
				(2.68)				(-0.31)
Nagelkerke's R^2	0.07	0.25	0.25	0.14	0.07	0.25	0.25	0.14
N	723	296	296	293	723	296	296	293

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5c-8c = Abitur; models 9c-12c = university transitions.

Table G: Sequential logit model of students' educational transitions (synthetic data, *PHEO* predictors)

	Model 5d	*Model 6d*	*Model 7d*	*Model 8d*	*Model 9d*	*Model 10d*	*Model 11d*	*Model 12d*
	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z
B	0.12	0.13	0.10	0.12	0.03	-0.11	-0.10	-0.10
	(0.86)	(0.92)	(0.62)	(0.76)	(0.11)	(-0.35)	(-0.29)	(-0.31)
-SD	-0.12	-0.14	-0.13	-0.13	0.46	0.47	0.43	0.40
	(-0.83)	(-0.89)	(-0.78)	(-0.74)	(1.38)	(1.26)	(1.13)	(1.01)
p(sd)	0.13	0.22	0.19	0.24	-0.01	0.23	0.11	0.16
	(0.90)	(1.38)	(1.16)	(1.40)	(-0.03)	(0.69)	(0.32)	(0.45)
C	0.30*	0.39**	0.44**	0.48**	0.18	0.36	0.27	0.28
	(2.12)	(2.62)	(2.82)	(2.98)	(0.59)	(1.13)	(0.81)	(0.83)
p(ep)	0.04	-0.04	-0.07	-0.04	0.26	0.45	-0.01	0.33
	(0.23)	(-0.21)	(-0.34)	(-0.17)	(0.60)	(0.90)	(-0.02)	(0.64)
par. occ. pres.	-0.01	-0.01	-0.01	-0.01	0.00	-0.00	-0.00	-0.00
	(-1.33)	(-1.24)	(-1.24)	(-1.10)	(0.13)	(-0.12)	(-0.21)	(-0.24)
par. educ.	0.23***	0.24***	0.22***	0.24***	0.05	0.07	-0.02	-0.00
	(3.92)	(3.78)	(3.33)	(3.52)	(0.36)	(0.48)	(-0.14)	(-0.03)
residuals (performance model)		-0.22				-0.55		
		(-1.01)				(-0.99)		
residuals (motivation model)			-0.28				-0.88	
			(-1.29)				(-1.53)	
residuals (full model)				-0.25				-0.63
				(-1.11)				(-1.12)
Nagelkerke's R^2	0.03	0.04	0.04	0.05	0.03	0.04	0.04	0.05
N	895	799	728	699	895	799	728	699

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5d-8d = Abitur; models 9d-12d = university transitions.

Table H: Sequential logit model of students' educational transitions (synthetic data, *PHEO* residuals and *CHiSP* SEU controls)

	Model 5e log-odds/z	*Model 6e* log-odds/z	*Model 7e* log-odds/z	*Model 8e* log-odds/z	*Model 9e* log-odds/z	*Model 10e* log-odds/z	*Model 11e* log-odds/z	*Model 12e* log-odds/z
B	0.08	0.15	0.10	0.17	-0.04	-0.11	-0.14	-0.16
	(0.54)	(0.86)	(0.54)	(0.89)	(-0.13)	(-0.33)	(-0.40)	(-0.46)
-SD	-0.15	-0.21	-0.14	-0.18	0.43	0.53	0.45	0.49
	(-0.86)	(-1.13)	(-0.69)	(-0.86)	(1.18)	(1.32)	(1.09)	(1.14)
p(sd)	0.15	0.23	0.12	0.14	-0.02	0.17	0.02	0.09
	(0.87)	(1.20)	(0.59)	(0.71)	(-0.06)	(0.48)	(0.06)	(0.24)
C	0.32	0.42*	0.48**	0.52**	0.25	0.38	0.32	0.35
	(1.95)	(2.38)	(2.61)	(2.76)	(0.78)	(1.11)	(0.90)	(0.98)
p(ep)	0.04	-0.08	-0.14	-0.15	0.58	0.69	0.24	0.56
	(0.20)	(-0.38)	(-0.60)	(-0.62)	(1.16)	(1.24)	(0.45)	(0.97)
B	0.23*	0.21*	0.32**	0.35**	-0.03	-0.03	0.08	0.13
	(2.38)	(2.02)	(2.94)	(3.12)	(-0.17)	(-0.13)	(0.39)	(0.59)
-SD	0.09	0.04	0.08	0.10	-0.18	-0.32	-0.31	-0.37
	(0.52)	(0.22)	(0.43)	(0.50)	(-0.56)	(-0.97)	(-0.88)	(-1.05)
p(sd)	-0.38*	-0.48*	-0.29	-0.36	-0.20	0.06	-0.05	0.05
	(-2.01)	(-2.39)	(-1.39)	(-1.65)	(-0.54)	(0.14)	(-0.12)	(0.12)
C	-0.09	-0.07	-0.11	-0.07	0.16	0.08	0.11	0.06
	(-0.55)	(-0.41)	(-0.60)	(-0.38)	(0.47)	(0.23)	(0.31)	(0.16)
p(ep)	1.59***	1.57***	1.57***	1.66***	-0.17	-0.04	0.10	0.26
	(4.13)	(3.84)	(3.58)	(3.58)	(-0.16)	(-0.04)	(0.09)	(0.24)
residuals (performance model)		-0.35				-0.27		
		(-1.32)				(-0.48)		
residuals (motivation model)			-0.37				-0.62	
			(-1.42)				(-1.07)	
residuals (full model)				-0.42				-0.38
				(-1.54)				(-0.67)
Nagelkerke's R^2	0.07	0.08	0.09	0.10	0.07	0.08	0.09	0.10
N	702	631	562	546	702	631	562	546

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5e-8e = Abitur; models 9e-12e = university transitions.

Table I: Sequential logit model of students' educational transitions (synthetic data, both *PHEO* and *CHiSP* residuals)

	Model 5f log-odds/z	*Model 6f* log-odds/z	*Model 7f* log-odds/z	*Model 8f* log-odds/z	*Model 9f* log-odds/z	*Model 10f* log-odds/z	*Model 11f* log-odds/z	*Model 12f* log-odds/z
B	0.13	0.15	0.20	0.40	0.07	-0.36	-0.33	-0.35
	(0.98)	(0.59)	(0.72)	(1.58)	(0.26)	(-0.80)	(-0.70)	(-0.75)
-SD	-0.13	-0.20	-0.50	-0.38	0.51	0.48	0.63	0.60
	(-0.98)	(-0.72)	(-1.75)	(-1.45)	(1.54)	(0.87)	(1.04)	(0.99)
p(sd)	0.13	0.19	0.30	0.27	0.04	0.10	0.22	0.16
	(0.89)	(0.65)	(1.01)	(0.99)	(0.14)	(0.21)	(0.42)	(0.31)
C	0.25	0.72**	0.66*	0.49	0.10	0.53	0.57	0.53
	(1.85)	(2.64)	(2.34)	(1.88)	(0.33)	(1.04)	(1.12)	(1.01)
p(ep)	0.04	-0.02	-0.19	-0.12	0.29	15.76	14.75	15.06
	(0.22)	(-0.06)	(-0.53)	(-0.36)	(0.66)	(0.01)	(0.02)	(0.02)
residuals (performance model)		-0.24				-0.34		
		(-0.61)				(-0.42)		
residuals (motivation model)			-0.45				-0.40	
			(-1.12)				(-0.49)	
residuals (full model)				-0.48				-0.42
				(-1.28)				(-0.52)
residuals (performance model)		2.50***				0.04		
		(9.14)				(0.06)		
residuals (motivation model)			2.43***				-0.07	
			(8.71)				(-0.11)	
residuals (full model)				1.28***				0.10
				(5.10)				(0.20)
Nagelkerke's R^2	0.01	0.34	0.34	0.15	0.01	0.34	0.34	0.15
N	945	351	321	303	945	351	321	303

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5f-8f = Abitur; models 9f-12f = university transitions.

Table J: Sequential logit model of students' educational transitions (synthetic data, both *PHEO* and *CHiSP* residuals and SEU controls)

	Model 5g log-odds/z	*Model 6g* log-odds/z	*Model 7g* log-odds/z	*Model 8g* log-odds/z	*Model 9g* log-odds/z	*Model 10g* log-odds/z	*Model 11g* log-odds/z	*Model 12g* log-odds/z
B	0.08	-0.02	-0.00	0.20	-0.04	-0.43	-0.42	-0.40
	(0.54)	(-0.06)	(-0.00)	(0.61)	(-0.13)	(-0.83)	(-0.78)	(-0.74)
-SD	-0.15	-0.52	-0.73*	-0.58	0.43	1.10	1.45	1.44
	(-0.86)	(-1.59)	(-2.10)	(-1.67)	(1.18)	(1.55)	(1.73)	(1.69)
p(sd)	0.15	0.13	0.16	-0.02	-0.02	0.10	0.11	0.10
	(0.87)	(0.38)	(0.45)	(-0.06)	(-0.06)	(0.17)	(0.19)	(0.16)
C	0.32	1.15***	1.22***	1.08**	0.25	0.75	0.73	0.71
	(1.95)	(3.46)	(3.43)	(3.14)	(0.78)	(1.32)	(1.25)	(1.20)
p(ep)	0.04	-0.03	-0.11	-0.19	0.58	15.97	15.38	15.80
	(0.20)	(-0.09)	(-0.26)	(-0.49)	(1.16)	(0.01)	(0.02)	(0.01)
B	0.23*	0.07	0.18	0.39	-0.03	-0.56	-0.40	-0.40
	(2.38)	(0.38)	(0.86)	(1.91)	(-0.17)	(-1.46)	(-0.99)	(-0.99)
-SD	0.09	0.07	0.26	0.37	-0.18	0.03	-0.05	-0.04
	(0.52)	(0.24)	(0.76)	(1.14)	(-0.56)	(0.06)	(-0.10)	(-0.08)
p(sd)	-0.38*	-0.29	-0.30	-0.32	-0.20	0.52	0.43	0.41
	(-2.01)	(-0.80)	(-0.78)	(-0.87)	(-0.54)	(0.74)	(0.60)	(0.56)
C	-0.09	0.26	0.29	0.20	0.16	-0.10	-0.34	-0.30
	(-0.55)	(0.82)	(0.83)	(0.61)	(0.47)	(-0.18)	(-0.60)	(-0.52)
p(ep)	1.59***	1.16	1.04	1.87*	-0.17	1.26	0.95	1.20
	(4.13)	(1.51)	(1.49)	(2.49)	(-0.16)	(0.73)	(0.60)	(0.76)
residuals (performance model)		-0.60				0.03		
		(-1.18)				(0.03)		
residuals (motivation model)			-0.66				-0.01	
			(-1.32)				(-0.01)	
residuals (full model)				-0.56				-0.05
				(-1.17)				(-0.05)
residuals (performance model)		1.91***				0.29		
		(5.69)				(0.33)		
residuals (motivation model)			1.85***				0.10	
			(5.23)				(0.12)	
residuals (full model)				0.52				0.19
				(1.63)				(0.32)
Nagelkerke's R^2	0.07	0.31	0.31	0.20	0.07	0.31	0.31	0.20
N	702	260	232	222	702	260	232	222

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 5g-8g = Abitur; models 9g-12g = university transitions.

Table K: Questions and categories of all variables taken from the 1980 British Cohort Study

Initial categories	**Concept**	**Indicator**
Students' performance	Average of student estimate of ability in math, literature and spelling	0 = "not so well" 1 = "well"
Students' motivation	Feel shy in front of teacher; feel foolish in front of peers; feel foolish with teacher (all student's answer)	0 = "yes" 1 = "don't know" 2 = "no"
	Continue training after leaving school (mother's answer)	0 = "no" 1 = "cannot say" 2 = "yes"
B	Type of training/education after school (mother's answer)	1 = "further" 2 = "college" 3 = "apprenticeship" 4 = "specific" 5 = "don't know"
-SD	Mother or father overconcerned? (teacher's answer)	0 = "does not apply" 1 = "yes applies"
p(SD)	Useless to try in school (student's answer)	0 = "yes" 1 = "don't know" 2 = "no"
	Studying for tests is waste of time (student's answer)	0 = "yes" 1 = "don't know" 2 = "no"
C	Total gross family income (parents' answer)	0 = "under 35pw" 1 = "35 - 49pw" 2 = "50 - 99pw" 3 = "100 - 149pw" 4 = "150 - 199pw" 5 = "200 - 249pw" 6 = "250 or more"
p(ep)	Low marks even though study hard	0 = "yes" 1 = "don't know" 2 = "no"

Table L: Questions and categories of all variables taken from the 1986 British Cohort Study

Teachers' expectations	Assessment of student's academic ability	0 = "bottom 5%" 1 = "well below aver 10%" 2 = "below average 20%" 3 = "Average 30%" 4 = "above average 20%" 5 = "well above aver 10%" 6 = "top 5%"
Students' performance	Average grade summer term exam (BCS1986)	0 = "fail" 1 = "CSE 4" 2 = "'O'lev E/'A' E/CSE 3" 3 = "'O'lev D/'A' D/CSE 2" 4 = "C/CSE 1" 5 = "'O'lev C/'A' " 6 = "'O'lev B/'A' B" 7 = "'O'lev A/'A' A"
Students' motivation	Quiet in classroom and get on with work	0 = "very true" 1 = "partly true" 2 = "not true at all"
	Think homework is a bore	0 = "very true" 1 = "partly true" 2 = "not true at all"
	I don't like school	0 = "very true" 1 = "partly true" 2 = "not true at all"
B	Job aspiration: Career in a profession (need a degree) [student's answer]	0 = "might do it" 1 = "joint first choice" 2 = "first choice"
	Leave reason: always taken for granted (student's answer)	0 = "not stated" 1 = "yes"
-SD	Satisfaction with child's school progress (mother's answer)	0 = "not satisfied" 1 = "can't say" 2 = "fairly satisfied" 3 = "very satisfied"
p(SD)	Any of your subjects useful in future? (student's answer)	0 = "many not useful" 1 = "some are useful" 2 = "all are useful"
	Feel school is largely a waste of time (student's answer)	0 = "very true" 1 = "partly true" 2 = "not true at all"
C	Fam troubled by finance hardshp past year (parents' answer)	0 = "no" 1 = "uncertain / don't know" 2 = "yes"
p(ep)	Leave reason: might be not bright enough (student's answer)	0 = "not stated" 1 = "yes"
	Find it difficult to keep mind on work	0 = "very true" 1 = "partly true" 2 = "not true at all"

Table M: Summary statistics of all indicators generated from the British Cohort Study

	count	mean	sd	min	max
Educational transitions (BCS combined)	11239	1.12	1.26	0.00	3.00
Teachers' expectations BCS 1980 dichotomized	16117	0.45	0.50	0.00	1.00
Residuals perf. (BCS 1980)	12491	0.00	0.44	-0.51	0.93
Residuals mot. (BCS 1980)	12382	0.00	0.46	-0.37	0.80
Residuals full (BCS 1980)	12219	-0.00	0.44	-0.52	1.03
Residuals perf. dichot. (BCS 1980)	16117	0.47	0.50	0.00	1.00
Residuals mot. dichot. (BCS 1980)	16117	0.47	0.50	0.00	1.00
Residuals full dichot. (BCS 1980)	16117	0.49	0.50	0.00	1.00
B (BCS 1980)	15571	0.14	0.35	0.00	1.00
-SD (BCS 1980)	12755	0.09	0.29	0.00	1.00
p(SD) (BCS 1980)	12510	0.71	0.45	0.00	1.00
C (BCS 1980)	16117	0.44	0.50	0.00	1.00
p(ep) (BCS 1980)	12967	0.54	0.50	0.00	1.00
Teachers' expectations BCS 1986 dichotomized	16117	0.88	0.32	0.00	1.00
Average grade summer term exam (BCS 1986)	16117	0.62	0.00	0.62	0.62
Residuals perf. (BCS 1986)	16117	-0.00	0.32	-0.88	0.12
Residuals mot. (BCS 1986)	6179	-0.00	0.42	-0.87	0.34
Residuals full (BCS 1986)	6179	-0.00	0.42	-0.87	0.34
Residuals perf. dichot. (BCS 1986)	16117	0.88	0.32	0.00	1.00
Residuals mot. dichot. (BCS 1986)	16117	0.91	0.28	0.00	1.00
Residuals full dichot. (BCS 1986)	16117	0.91	0.28	0.00	1.00
B (BCS 1986)	6417	0.20	0.40	0.00	1.00
SD (BCS 1986)	16117	0.07	0.26	0.00	1.00
p(SD) (BCS 1986)	6281	0.74	0.44	0.00	1.00
C (BCS 1986)	16117	0.53	0.50	0.00	1.00
p(ep) (BCS 1986)	6324	0.53	0.50	0.00	1.00

Table N: Sequential logit model of students' educational transitions (British Cohort Study, primary school residuals and SEU controls)

	Model 1h log-odds/z	*Model 2h* log-odds/z	*Model 3h* log-odds/z	*Model 4h* log-odds/z	*Model 5h* log-odds/z	*Model 6h* log-odds/z	*Model 7h* log-odds/z	*Model 8h* log-odds/z	*Model 9h* log-odds/z	*Model 10h* log-odds/z	*Model 11h* log-odds/z	*Model 12h* log-odds/z
B	0.98***	0.81***	0.81***	0.83***	1.29***	1.08***	1.07***	1.10***	0.16	0.16	0.17	0.17
	(12.54)	(10.08)	(10.02)	(10.36)	(13.02)	(10.59)	(10.50)	(10.75)	(1.16)	(1.15)	(1.23)	(1.21)
-SD	0.05	0.05	0.04	0.04	0.07	0.06	0.07	0.06	-0.08	-0.08	-0.08	-0.08
	(0.54)	(0.48)	(0.47)	(0.47)	(0.52)	(0.44)	(0.51)	(0.47)	(-0.40)	(-0.40)	(-0.41)	(-0.40)
p(sd)	0.28***	0.21**	0.21***	0.23***	0.39***	0.29**	0.30**	0.32***	-0.61**	-0.61**	-0.61**	-0.61**
	(4.38)	(3.28)	(3.30)	(3.57)	(4.17)	(3.04)	(3.11)	(3.38)	(-2.95)	(-2.94)	(-2.93)	(-2.93)
C	0.42***	0.36***	0.36***	0.37***	0.56***	0.48***	0.47***	0.49***	0.39**	0.39**	0.39**	0.39**
	(7.77)	(6.52)	(6.47)	(6.67)	(7.56)	(6.29)	(6.22)	(6.44)	(3.12)	(3.11)	(3.15)	(3.14)
p(ep)	0.69***	0.58***	0.58***	0.62***	0.63***	0.45***	0.45***	0.49***	-0.15	-0.15	-0.14	-0.14
	(10.68)	(8.72)	(8.81)	(9.41)	(6.29)	(4.38)	(4.41)	(4.77)	(-0.72)	(-0.70)	(-0.64)	(-0.66)
residuals (performance model)		0.70***				0.95***				-0.01		
		(11.76)				(12.07)				(-0.08)		
residuals (motivation model)			0.70***				0.95***				-0.06	
			(11.87)				(12.17)				(-0.45)	
residuals (full model)				0.62***				0.88***				-0.05
				(10.83)				(11.41)				(-0.35)
Nagelkerke's R^2	0.15	0.20	0.20	0.19	0.15	0.20	0.20	0.19	0.15	0.20	0.20	0.19
N	6386	6386	6386	6386	6386	6386	6386	6386	6386	6386	6386	6386

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 1h-4h= O level 5h-8h = A level; models 9h-12h = university transitions.

Table O: Sequential logit model of students' educational transitions (British Cohort Study, secondary school residuals and SEU controls)

	Model 1i	*Model 2i*	*Model 3i*	*Model 4i*	*Model 5i*	*Model 6i*	*Model 7i*	*Model 8i*	*Model 9i*	*Model 10i*	*Model 11i*	*Model 12i*
	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z	log-odds/z
B	1.58***	1.47***	1.47***	1.47***	2.04***	1.90***	1.90***	1.90***	0.83***	0.84***	0.84***	0.84***
	(15.49)	(14.21)	(14.27)	(14.27)	(16.14)	(14.90)	(14.94)	(14.94)	(5.88)	(5.93)	(5.94)	(5.94)
-SD	-0.87***	-0.83***	-0.83***	-0.83***	-1.10***	-0.94***	-0.95***	-0.95***	0.51	0.50	0.50	0.50
	(-7.96)	(-7.47)	(-7.47)	(-7.47)	(-6.05)	(-5.10)	(-5.11)	(-5.11)	(1.26)	(1.23)	(1.22)	(1.22)
p(sd)	0.27***	0.25***	0.25***	0.25***	0.63***	0.60***	0.61***	0.61***	-0.11	-0.10	-0.10	-0.10
	(3.89)	(3.53)	(3.54)	(3.54)	(6.59)	(6.20)	(6.25)	(6.25)	(-0.65)	(-0.63)	(-0.63)	(-0.63)
C	-0.44***	-0.39***	-0.39***	-0.39***	-0.21*	-0.15	-0.15	-0.15	0.08	0.07	0.07	0.07
	(-6.60)	(-5.74)	(-5.81)	(-5.81)	(-2.32)	(-1.54)	(-1.57)	(-1.57)	(0.53)	(0.48)	(0.48)	(0.48)
p(ep)	-0.18**	-0.20**	-0.20**	-0.20**	-0.95***	-0.97***	-0.97***	-0.97***	0.39**	0.39**	0.39**	0.39**
	(-2.93)	(-3.24)	(-3.24)	(-3.24)	(-11.47)	(-11.45)	(-11.45)	(-11.45)	(3.00)	(2.99)	(2.99)	(2.99)
residuals (performance model)		0.74***				1.10***				-0.17		
		(10.25)				(9.83)				(-0.75)		
residuals (motivation model)			0.73***				1.09***				-0.19	
			(10.05)				(9.75)				(-0.86)	
residuals (full model)				0.73***				1.09***				-0.19
				(10.05)				(9.75)				(-0.86)
Nagelkerke's R^2	0.23	0.27	0.27	0.27	0.23	0.27	0.27	0.27	0.23	0.27	0.27	0.27
N	5042	5042	5042	5042	5042	5042	5042	5042	5042	5042	5042	5042

All coefficients are unstandardized log-odds. 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; C: expected costs; p_{ep}: subjective expected probability of educational success.
Educational transitions: Models 1i-4i= O level 5i-8i = A level; models 9i-12i = university transitions.

Curriculum Vitae

Dominik Becker, M.A.

DATE OF BIRTH May 16, 1982

PLACE OF BIRTH Cologne, Germany

CONTACT INFORMATION

Technical University of Dortmund
Institute for School Development Research
Martin-Schmeißer-Weg 13
D - 44227 Dortmund

Phone: +49 (0)231 755-6203
E-mail: becker@ifs.tu-dortmund.de
Web: http://www.ifs-dortmund.de
Fax: +49 (0)231 755-7809

POSITION

Research assistant (full position) in the quantitative evaluation of the research project "Ganz In – Mit Ganztag mehr Zukunft. Das neue Ganztagsgymnasium NRW ["Ganz In – All Day-Schools for a Brighter Future. The New All-Day Secondary School in NRW"]

Main activities and responsibilities:

- Quantitative assessment of project schools' educational effectiveness
- Design and implementation of quantitative surveys in 31 upper secondary schools ($N \approx 7.000$ students and parents in both 5^{th} and 7^{th} grade, $N \approx 2.000$ teachers)
- Co-supervision of Bachelor, Diploma/Master, and Doctoral theses
- Project coordination
- Teaching (statistical methods, philosophy of the social sciences, social action theory)

EDUCATION

October 2008 – September 2011

Ph.D. fellow at the Cologne Graduate School in Management, Economics and Social Sciences

- Thesis title: *Pygmalion's Long Shadow – Determinants and Outcomes of Teachers' Evaluations* (defended December 5, 2012; *summa cum laude*)
- Supervisors: Professor em. Heiner Meulemann, Professor Rolf Becker
- Area of study: Sociology of Education

October 2005 – February 2006

Semester abroad at the University of Fribourg (funded by *ERASMUS* and *Studienstiftung des Deutschen Volkes*)

October 2002 – February 2008

Magister Artium, Sociology, Philosophy and German Literature

- Thesis Topic: *Politische Homogenität oder Heterogenität der Leserschaft von Qualitätszeitungen? Eine Sekundäranalyse der Media-Analyse-Daten 2005 [Political homogeneity or heterogeneity of quality paper readership? A secondary analysis of the German Media Analysis data 2005]*
- Supervisors: Professor em. Heiner Meulemann, Professor em. Jürgen Friedrichs
- Area of Study: Communication science / media sociology
- Overall grade: *very good* (1.2)

Research Experience

October 2011 –

Research assistant (full position) at the Institute for School Development Research of the Technical University of Dortmund in the quantitative evaluation of the research project "Ganz In – Mit Ganztag mehr Zukunft. Das neue Ganztagsgymnasium NRW [*"Ganz In – All Day-Schools for a Brighter Future. The New All-Day Secondary School in NRW"*], a joint project initiated by Stiftung Mercator, Institut für Schulentwicklungsforschung Dortmund – representing the three universities in the Ruhr region – and the Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen [the Ministery for School and Further Educational of the German Federal State of North Rhine-Westphalia].

April – September 2008

Research assistant at the Institut für Berufs- und Weiterbildung (University of Duisburg-Essen). Cooperation in the research project *Lehrerurteile im Lebenslauf* [Teacher Evaluations in the Life Course], funded by the German Research Association (DFG). Research Report: Birkelbach, K. and R. Dobischat (eds., 2010): Lehrerurteile im Lebensverlauf. Abschlussbericht an die DFG zum Projekt DO 325 4/1. Essen: University of Duisburg-Essen.

January 2007 - March 2008

Student assistant at the Medienwissenschaftliches Lehr- und Forschungszentrum (University of Cologne). Cooperation in the research project *Elektronischer Datenreport Media Analyse: Lesen von Zeitungen und Zeitschriften in Deutschland von 1954 bis 2002* [Electronic Data Report Media Analysis: Usage of Newspapers and Magazines in Germany, 1954-2002], funded by the German Federal Ministry of Education and Research (BMBF).

October 2004 – August 2005

Universitary research practical at the Research Institute for Sociology, University of Cologne (supervisor: Prof. Michael Wagner). Research report: Becker, D. and St. Lang (2005): Determinanten von Schulverweigerung im Internationalen Vergleich [A Cross-National Comparison of Determinants of School Absenteeism]. Unpublished Manuscript, University of Cologne.

Professional Experience

October 2001 - June 2002

Temporarily employed at *Axa Insurance Company.*

- Field: Internal preparation of Holocaust compensations

September 2001

Civilian service at the *Arbeiter-Samariter-Bund* (ASB) Cologne.

- Field: Assistance with transport

Stipends & Awards

July 2011

Travel allowance by the *German Academic Exchange Service (DAAD)* for the conference talks "Analyzing social and media change with cross-classified random-effects APC models", "Creating artificial student cohorts by statistical matching of distinct educational data sources", and "Analyzing contextual-level outcomes in multilevel models", presented at at the 4th Conference of the *European Survey Research Association (ESRA)*, Lausanne, July 19, 2011

September 2010

Best paper proposal award by the *Section on Methods of the German Association of Publication and Communication Research (DGPuK)* for the conference talk “Nimmt die Lesewahrscheinlichkeit von Qualitätszeitungen mit der Bildungsexpansion zu? Eine Mehrebenen-Anayse der Media-Analyse-Daten 1977-2006” [Does educational expansion foster the probability of reading quality papers? A multilevel analysis of the German Media Analysis data 1977-2005] (September 24, 2010, Vienna; with J. Hagenah)

August 2010

Travel allowance by the *German Academic Exchange Service (DAAD)* for the conference talk “The impact of teachers’ expectations on educational opportunities in the life course - An ignored inequality?” (presented at at the XVII ISA World Congress of Sociology, July 17, 2010, Gothenburg)

Mai 2010

Travel award by the *Modern Austrian Literature and Culture Association (MALCA)* for the conference talk “Peter Handkes ’Bildverlust’ im Kontext einer systemtheoretischen Interpretation seines frühen Motivs der ’Innenwelt der Außenwelt der Innenwelt”’ (Mai 24, 2010, Vienna)

October 2008 – September 2011

PhD scholarship by the *Cologne Graduate School in Management, Economics and Social Sciences* (funded by the federal government of North Rhine-Westphalia)

August 2006

Polish language course at Jagiellonen University of Krakow, funded by the *Studienstiftung des deutschen Volkes* [German National Academic Foundation]

October 2005 – February 2006

Foreign exchange scholarship of the European Union (*Erasmus grant*). Additional foreign exchange scholarship by the *Studienstiftung des deutschen Volkes* [German National Academic Foundation]

March 2005 – January 2008

Full scholarship of the *Studienstiftung des deutschen Volkes* [German National Academic Foundation]

Primary Research Interests

Inequalities in educational opportunities and school effectiveness
Rational choice theory and analytical sociology
Statistical methods in social sciences
Political sociology, values and attitudes
Communication research
Philosophy of the social sciences

Teaching Experience

Summer term 2013

Research practical (master level): Secondary analysis of the “Ganz In” data (with S. Lindemann, J. Schwanenberg & D. Winkelsett).

Winter term 2012

PhD course “Wissenschaftstheorie” [Philosophy of the social sciences]. One-day block course taught in the context of the mandatory PhD program of the Institute for School Development Research of the Technical University of Dortmund (with S. Lindemann).

PhD course “Qualitative und quantitative Methoden, Triangulation” [Qualitative and quantitative methods, mixed method design]. Three-day block course taught in the context of the mandatory PhD program of the Institute for School Development Research of the Technical University of Dortmund (with A. Walzebug).

Master-level course “Wissenschafts- und handlungstheoretische Grundlagen empirischer Bildungsforschung” [Philosophy of the social sciences, social action theory, and empirical educational research]. Weekly seminar.

Summer term 2012

PhD course “Von der Forschungsfrage zur Statistik – Workshop *quantitative Methoden*” [From research questions to coefficients: Workshop *quantitative methods of analysis*]. Block course taught at the official inception of the research forum at the Dortmunder Kompetenzzentrum für Lehrerbildung und Lehr-/Lernforschung of the Technical University of Dortmund (with K. Drossel).

Summer term 2008

Bachelor course “Raumsoziologie” [Sociology of Space]. Block course taught at the Institute for urban district development, social casework and welfare mentoring of the University of Duisburg-Essen

Membership in Associations

ISA Research Committee on Social Stratification and Mobility (RC28 of the International Sociological Association, ISA)

European Survey Research Association (ESRA)

ISA Research Committee on Logic and Methodology (RC33 of the International Sociological Association, ISA)

Association for Educational Assessment (AEA) Europe

American Educational Research Association (AERA), Division D “Measurement and Research Methodology”

Skills

Languages

- German: Native
- English: Fluent
- French: Advanced
- Polish: Basic – intermediate

IT

- Operating systems: Windows, Linux, OSX
- Text and office processing: MS Office, OpenOffice, LaTeX
- Statistical packages: `Stata`, `SAS`, `SPSS`, `R`, `AMOS`, `Mplus`, `Conquest`
- Referencing software: Mendeley, Zotero, Citavi

PUBLICATIONS

Becker, D. (forthcoming): The impact of teachers' expectations on students' educational inequality in the life course – An empirical test of a subjective expected utility explanation. Forthcoming in *Rationality & Society.*

Becker, D. & K. Birkelbach (forthcoming): Lehrer als 'Gatekeeper'? Eine theoretische Annäherung an Determinanten und Folgen von Lehrererwartungen [Teachers as 'gate-keepers'? A theoretical approximation towards determinants and outcomes of teachers' evaluations]. To appear in: Schulze, A. & R. Becker (eds.): Bildungskontexte. Strukturelle Vorraussetzungen und Ursachen ungleicher Bildungschancen. Wiesbaden: VS Verlag.

Hagenah, J., D. Gilles & D. Becker (2013): Überregionale Tageszeitungen: Eine aussterbende Informationsressource? [Nationwide Newspapers: Information resources in danger of extinction?] In: Ohly, H.-P. (Ed.): Wissen – Wissenschaft – Organisation. Proceedings der 12. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation, Bonn, 19. bis 21. Oktober 2009. Würzburg: Ergon Verlag, pp. 98-105.

Schwanenberg, J., N. McElvany, D. Becker & N. Pfuhl (2013): Elternpartizipation an Grundschulen unter Berücksichtigung des sozialen und kulturellen Familienhintergrunds [Parental involvement in the context of social and cultural family backgrounds]. In: McElvany, N., M. M. Gebauer, W. Bos & H. G. Holtappels (Eds.): Sprachliche, kulturelle und soziale Heterogenität in der Schule als Herausforderung und Chance der Schulentwicklung. IfS-Jahrbuch der Schulentwicklung, Bd. 17. Weinheim: Juventa, pp. 150-180.

Becker, D. & J. Hagenah (2012): Zur Anwendbarkeit von Mehrebenenmodellen bei der integrativen Untersuchung von Alters-, Perioden- und Kohorteneffekten Exemplarische Untersuchungen mit den Media-Analyse-Daten 1977-2006 [The use of multilevel models for simultaneous analyses of age, period and cohort effects]. In: Stark, B. and M. Magin (eds.): *Methodische Herausforderungen komparativer Forschungsansätze.* Köln: Herbert von Halem, pp. 211-238.

Hagenah, J. & D. Becker (2012): Über Nutzungswahrscheinlichkeiten und fusionierte Datensätze. Werbeträgerkontaktchancen als intermedial vergleichbare Medienwährungen in Deutschland [Usage probabilities and data fusion. Media contact probabilities as an intermedial 'currency' in Germany]. In: Hagenah, J. & H. Meulemann (eds.): *Mediatisierung der Gesellschaft?* Münster: Lit Verlag, pp. 99-117.

Becker, D. (2011): Konsonante oder dissonante Nutzung von Qualitätszeitungen? Eine Sekundäranalyse der Media-Analyse-Daten 2005 [Consonant or dissonant usage of quality papers? A secondary analysis of the German Media Analysis data 2005]. In: Haschke, J.F. & A. Moser (eds.): *Politik – Deutsch, Deutsch – Politik: Aktuelle Trends und Forschungsergebnisse.* Beiträge zur 6. Tagung des Düsseldorfer Forum Politische Kommunikation (DFPK). Berlin: Frank und Timme, pp. 199-222.

Becker, D. & K. Birkelbach (2010): Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond- oder Reflected-Glory-Effekt? Eine Mehrebenen-Analyse von Lehrerurteilen [Intelligence and school achievement as contextual-level predictors: Big-Fish-Little-Pond- or Reflected-Glory-Effect? A multilevel analysis of teachers' evaluations]. In: Becker, T., K. Birkelbach, J. Hagenah & U.

Rosar (eds.): *Komparative empirische Sozialforschung.* Wiesbaden: VS Verlag, pp. 113-141.

Becker, D. (2009): Dahinter steckt meistens ein kluger Kopf. Mehr und dissonantere Leitmedienlektüre infolge der Bildungsexpansion? [Did educational expansion lead to more frequent and more dissonant usage of quality papers?] In: Müller, D., A. Ligensa & P. Gendola (eds.): *Alte und neue Leitmedien*, Vol. 1. Bielefeld: Transcript Verlag, pp. 199-224.

Becker, D. (2008): Qualitätszeitungen – regionale Bindungseffekte trotz überregionaler Ausrichtung? Ein Längsschnittvergleich der Media-Analyse-Daten 1985, 1995 und 2005 [Quality papers – regional recruitment of readership in face of nationwide demand?]. In: Hagenah, J. & H. Meulemann (eds.): *Alte und neue Medien. Zum Wandel der Medienpublika in Deutschland seit den 1950er Jahren.* Münster: Lit Verlag, pp. 61-80.

Full Conference Papers

Becker, D., A.S. Willems, R. Springer & W. Bos (2013): A doubly-latent model of class-level social status and achievement on students' academic self-concept. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), April 27, 2013, San Francisco.

Becker, D., J. Schwanenberg & D. Winkelsett (2012): School resources, inequality in educational opportunities, and school effectiveness – Multilevel analyses of a German upper-track secondary school sample. Paper presented at the 13^{th} conference of the Association for Educational Assessment (AEA) Europe, November 9, 2012, Berlin.

Becker, D., K. Drossel, J. Schwanenberg, N. Pfuhl & H. Wendt (2012): Comparing IRT, CFA and LCA for Assessing the Social Composition of School Classes. Paper presented at the *RC33 Eighth International Conference on Social Science Methodology*, July 10, 2012, Sydney.

Schwanenberg, J., D. Becker, N. McElvany, N. Pfuhl & R. Springer (2012): Parents' engagement in the school context – An analysis of areas of engagement, differences in intensity, and relevant motivational aspects within an expectancy-value framework. Paper presented at the 2012 Annual Meeting of the *American Educational Research Association*, April 16, 2012, Vancouver (paper presented by N. McElvany).

Becker, D. (2011): Creating artificial student cohorts by statistical matching of distinct educational data sources. Paper presented at the fourth Conference of the *European Survey Research Association (ESRA)*, July 19, 2011, Lausanne.

Becker, D. (2011): Does the effect of teachers' expectations on students' educational opportunities decrease over educational transitions? A statistical matching approach. Paper presented at the *2011 Spring Meeting of the ISA RC28 on Social Stratification and Mobility*, April 14, 2011, Essex.

Becker, D., W. Breustedt & C. Zuber (2011): Citizens, elites, and the democratic game. A two-step HLM analysis of democratic stability. Paper presented at the ECPR-IPSA Joint Conference *Whatever happened to North-South*? February 18, 2011, Sao Paulo (Paper presented by C. Zuber).

Manuscripts in Preparation

Becker, D., A.S. Willems & W. Bos (accepted): Referenzgruppeneffekte im Mehrebenen-Strukturgleichungsmodell [Reference-group effects in multilevel structural equation models]. To appear in: Holtappels, H.-G. & M. Pfeifer (eds.): Jahrbuch der Schulentwicklung 2014 (Jahrbuch 18). Weinheim: Juventa.

Becker, D., K. Drossel, J. Schwanenberg, H. Wendt & W. Bos (submitted): Der Sozialindex: Theoretische Fundierung und forschungspraktische Relevanz für die

Erfassung der Schülerkomposition von Gymnasien [The social composition index: theoretical foundation and practical relevance for German upper-secondary schools' social composition]. To appear in: Grooth-Wilken, B., K. Isaac and J.-P. Schräpler (eds.): "Sozialindex – Modelle und Anwendungsgebiete" (working title). Münster: Waxmann.

Becker, D., R. Strietholt, J. Schwanenberg & W. Bos (submitted): Ganztags-schule – ganz gut? [All-day school – a success story?] To appear in: Schulte, D. (ed.): Zukunft des Lernens. Paderborn: Wilhelm Fink Verlag.

Working Papers

Becker, D. & K. Birkelbach (2011): Teachers' evaluations and the definition of the situation in the classroom. CGS Working Paper Series No. 02-04. Cologne Graduate School in Management, Economics and Social Sciences.

Becker, D. (2010): The impact of teachers' expectations on students' educational opportunities in the life course. CGS Working Paper Series No. 01-07. Cologne Graduate School in Management, Economics and Social Sciences.

Becker, D. (2007): Politische Homogenität oder Heterogenität der Leserschaft von Qualitätszeitungen? Eine Sekundäranalyse der Media-Analyse-Daten 2005 [Political homogeneity or heterogeneity of quality paper readership? A cross-sectional analysis of the German Media Analysis data 2005]. Unpublished Magister Thesis, Seminar für Soziologie, University of Cologne.

Becker, D. & St. Lang (2005): Determinanten von Schulverweigerung im Internationalen Vergleich [A Cross-National Comparison of Determinants of School Absenteeism]. Unpublished research report, University of Cologne.

Symposia and conference sessions

Becker, D. (2013): Generalized Latent Variable Modeling. Conference session organized for the 5th *European Survey Research Association (ESRA)* conference, July 18, 2013, Ljubljana.

Becker, D. (2013): Substantive studies using cross-national data. Conference session accepted at the 5th *European Survey Research Association (ESRA)* conference, July 16, 2013, Ljubljana.

Becker, D. & K. Drossel (2012): Motivation in the teaching profession and beyond. Symposium organized for the 2012 *International Conference of Motivation (ICM)* of the Special Interest Group on Motivation and Emotion (SIG 8) of the European Association for Research on Learning and Instruction (EARLI), August 28, 2012, Frankfurt a.M.

Accepted Abstracts

Becker, D., A.S. Willems & W. Bos (accepted): Schuleffektivität und nicht-kognitive Outcomes - Neue Erkenntnisse aus einer aktuellen Panelstudie [School effectiveness and non-cognitive outcomes - new evidence from current panel data]. Abstract accepted at the 78th *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 25-27, Dortmund.

Schwanenberg, J., D. Becker, W. Bos & N. McElvany (accepted): Warum engagieren sich Eltern in der Schule ihres Kindes und zu Hause? Eine Analyse der Motive unter Anwendung des Erwartungs-Wert-Modells [Why do parents become involved at school and at home? An expectancy-value analysis of parental motives]. Abstract accepted at the 78th *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 25-27, Dortmund.

Winkelsett, D., A.S. Willems, D. Becker & W. Bos (accepted): Der Big-Fish-Little-Pond-Effekt: Moderation durch Dimensionen des sozialen Selbstkonzepts [The Big Fish Little Pond effect: Moderation by dimensions of students' social self-concept]. Abstract accepted at the 78th *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 25-27, Dortmund.

Becker, D., A.S. Willems, R. Springer & W. Bos (accepted): Analyzing reference-group effects in secondary science education by using multilevel latent models. Abstract accepted at the 15^{th} Biennial Conference of the *European Association for Research on Learning and Instruction (EARLI)*, August 27-31, 2013, Munich.

CONFERENCE TALKS

Becker, D. & Schwanenberg, J. (2013): Multiple-group comparison and differential item functioning: Comparing two methods for assessing measurement invariance. Talk presented at the 5^{th} *European Survey Research Association (ESRA)* conference, July 18, 2013, Ljubljana.

Becker, D., T. Beckers, K. Birkelbach & U. Tranow (2013): The frogs and props of the social sciences: A plea for the advancement of mechanism-type explanations in (multilevel) educational and survey research. Talk presented at the 5*th European Survey Research Association (ESRA)* conference, July 18, 2013, Ljubljana.

Becker, D., A.S. Willems, R. Springer & W. Bos (2013): Big-Fish-Little-Pond oder Reflected-Glory-Effekt? Referenzgruppeneffekte von Leistungskontext und sozialer Komposition auf das akademische Selbstkonzept im Mehrebenenstrukturgleichungsmodell [Big-Fish-Little-Pond or Reflected-Glory Effect? Reference-group effects of class-level achievement and social composition on academic self-concept in a multilevel-SEM framework]. Talk presented at the first conference of the *Gesellschaft für Empirische Bildungsforschung (GEBF)*, March 11, 2013, Kiel.

Becker, D. & J. Schwanenberg (2012): Multiple Group Comparisons vs. Differential Item Functioning: Ein Vergleich von zwei Verfahren zur Überprüfung von Messinvarianz am Beispiel Elternpartizipation [Multiple Group Comparisons vs. Differential Item Functioning: A review of two approaches assessing measurement invariance with an application to parent involvement]. Talk presented at the 77^{th} *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 10, 2012, Bielefeld.

Becker, D. & B. Glesemann (2012): Der Zusammenhang zwischen schulischer Komposition, dem Einsatz von Instrumenten individueller Förderung und Chancen-(un)gleichheit [The relationship between school composition, students' individual promotion, and educational (in)equality]. Talk presented at the 77^{th} *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 11, 2012, Bielefeld.

Lindemann, S., D. Becker & J. Schröder (2012): Individuelle Förderung, akademisches Selbstkonzept und Schülerleistungen in Abhängigkeit von der Prozessqualität des gebundenen Ganztags [Individual promotion, academic self-concept and student achievement in context of teaching quality in full-day education]. Talk presented at the 77^{th} *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 11, 2012, Bielefeld (presented by S. Lindemann and J. Schröder).

Winkelsett, D. & D. Becker (2012): Kooperation zwischen Lehrkräften und weiterem pädagogisch tätigem Personal: Schulstrukturelle Determinanten und Schülerleistungseffekte [Cooperation between teachers and school's pedagogic staff: Structural determinants on school level and student achievement outcomes]. Talk presented at the 77^{th} *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 11, 2012, Bielefeld.

Drossel, K., J. Schwanenberg & D. Becker (2012): Motivational prerequisites of teacher cooperation. Talk presented at the 2012 *International Conference of Motivation (ICM)* of the Special Interest Group on Motivation and Emotion (SIG 8) of the European Association for Research on Learning and Instruction (EARLI), August 28, 2012, Frankfurt a.M.

Schwanenberg, J., N. McElvancy, D. Becker, N. Pfuhl & W. Bos (2012): Zusammenarbeit von Eltern und Schule – Formen und Umfang [An analysis of both forms and extent of parent-teacher collaboration]. Talk presented at the 23. Kongress der *Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)*, March 12, 2012, Osnabrück.

Becker, D. & J. Hagenah (2011): Analyzing social and media change with cross-classified random-effects APC models. Talk presented at the fourth Conference of the *European Survey Research Association (ESRA)*, July 22, 2011, Lausanne.

Becker, D., W. Breustedt & C. Zuber (2011): Analyzing contextual-level outcomes in multilevel models. Talk presented at the fourth Conference of the *European Survey Research Association (ESRA)*, July 21^{st}, 2011, Lausanne.

Hagenah, J. & D. Becker (2010): Über Nutzungswahrscheinlichkeiten und fusionierte Datensätze. Werbeträgerkontaktchancen als intermedial vergleichbare Medienwährungen in Deutschland [Usage probabilities and data fusion. [Media contact probabilities as an intermedial 'currency' in Germany]. Talk presented at the 3^{rd} *MLFZ Workshop "Mediatisierung der Gesellschaft?"*, November 19, Cologne.

Becker, D. & J. Hagenah (2010): Nimmt die Lesewahrscheinlichkeit von Qualitätszeitungen mit der Bildungsexpansion zu? Eine Mehrebenen-Anayse der Media-Analyse-Daten 1977-2006 [Does educational expansion foster the probability of reading quality papers? A multilevel analysis of the German Media Analysis data 1977-2005]. Talk presented at the 3^{rd} *MLFZ Workshop "Mediatisierung der Gesellschaft?"*, November 19, Cologne.

Becker, D. (2010): Pornografisierung, Säkularisierung und Wertewandel. Eine Mehrebenen-Analyse der Media-Analyse-Daten 1954-2006 [Pornographization, secularization and value change. A multilevel analysis of the German media analysis data 1954-2006]. Talk presented at the *Joint Conference of the German Association of Media Pedagogy and Culture of Communication (GMK) and the German Association of Publication and Communication Research (DGPuK) on 'Pornografisierung von Gesellschaft!?'*, October 29, 2010, Cologne.

Becker, D. & J. Hagenah (2010): Nimmt die Lesewahrscheinlichkeit von Qualitätszeitungen mit der Bildungsexpansion zu? Eine Mehrebenen-Anayse der Media-Analyse-Daten 1977-2006 [Does educational expansion foster the probability of reading quality papers? A multilevel analysis of the German media analysis data 1977-2005]. Talk presented at the 12^{th} *Annual Meeting of the section on methods of the German Association of Publication and Communication Research (DG-PuK)*, September 24, 2010, Vienna.

Becker, D. & K. Birkelbach (2010): The impact of teachers' expectations on students' educational opportunities in the life course. Talk presented at the *XVII ISA World Congress of Sociology*, July 17, 2010, Gothenburg.

Becker, D. (2010): The impact of teachers' expectations on educational opportunities in the life course - An ignored inequality? Talk presented at the international conference *Higher education and beyond – Inequalities regarding entrance to higher education and educational credentials*, July 6, 2010, Monte Verita.

Becker, D. (2010): Peter Handkes 'Bildverlust' im Kontext einer systemtheoretischen Interpretation seines frühen Motivs der 'Innenwelt der Außenwelt der Innenwelt'. Talk presented at the *Annual Meeting of the Modern Austrian Literature and Culture Association*, Mai 24, 2010, Vienna.

Becker, D. (2010): Politische Homogenität oder Heterogenität der Leserschaft von Qualitätszeitungen? [Political homogeneity or heterogeneity of quality paper readership?] Talk presented at the *6. Düsseldorfer Forum Politische Kommunikation*, April 10, 2010, Schloss Mickeln (Düsseldorf).

Becker, D. & K. Birkelbach (2009): Intelligenz und Schulleistung als Kontextmerkmale: Big-Fish-Little-Pond- oder Reflected-Glory-Effekt? [Intelligence and school achievement as contextual-level predictors: Big-Fish-Little-Pond- or Reflected-Glory-Effect?] Talk presented at the Symposium *Komparative empirische Sozialforschung*, October 31, 2009, Cologne.

Hagenah, J., Becker, D. & Gilles, D. (2009): Überregionale Tageszeitungen: Eine aussterbende Informationsressource? [Nationwide Newspapers: Information resources in danger of extinction?] Talk presented at the *12th Conference of the International Society for Knowledge Organization (ISKO)*, October 21, 2009, Bonn.

Becker, D. (2009): Referenzgruppeneffekte als Determinanten des Bildungserfolgs: Ein Mehrebenen-Pfadmodell auf Basis des Kölner Gymnasiasten-Panels 1969/1970, 1985 & 1997 [Reference group effects as determinants of educational success: Evidence from a multilevel path analysis]. Talk presented at the 73^{th} *meeting of the German Association of Empirical Pedagogic Research (AEPF)*, September 29, 2009, Bochum.

Hagenah, J., D. Becker & M. Ehrenberg (2009): Werbeträgerkontaktchancen als intermediale Medienwährung [Media contact probabilities - an intermedial 'currency']. Talk presented at the Conference "State of the Art und Zukunft der Media- und Werbungsforschung", *Annual Meeting of the Swiss Association of Communication and Media Research (SGKM)*, March 26, 2009, Zürich.

Becker, D., J. Hagenah & M. Ehrenberg (2008): Von der Nutzungshäufigkeit zur Kontaktwahrscheinlichkeit. Nutzungschancen als dominierende Kennziffern in der Media- bzw. Werbeplanungspraxis im Validitätstest [From usage frequencies to contact probabilities. A validity test of convenient parameters in media and advertisement research]. Talk presented at the *Annual Meeting of the section on methods of the German Association of Publication and Communication Research (DGPuK)*, September 18, 2008, Bad Tölz.

Becker, D. (2007): Konsonante und dissonante Leitmediennutzung am Beispiel von Qualitätszeitungen. [Consonant and dissonant usage of quality papers]. Talk presented at the *Second MLFZ-Workshop "Alte und Neue Medien - Zum Wandel der Medienpublika in Deutschland seit den 1950er Jahren"*, Medienwissenschaftliches Lehr- und Forschungszentrum, October 5, 2007, Cologne.