

Bayer, Michael; Zinn, Sabine; Rüdiger, Christin

Article — Published Version

Grading in Secondary Schools in Germany – The Impact of Social Origin and Gender

International Journal of Educational Research Open

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Bayer, Michael; Zinn, Sabine; Rüdiger, Christin (2021) : Grading in Secondary Schools in Germany – The Impact of Social Origin and Gender, International Journal of Educational Research Open, ISSN 2666-3740, Elsevier BV, Amsterdam, Vol. 2, <https://doi.org/10.1016/j.ijedro.2021.100101> , <https://www.sciencedirect.com/science/article/pii/S2666374021000716?via%3Dihub>

This Version is available at:

<https://hdl.handle.net/10419/266368>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc-nd/4.0/>



Grading in Secondary Schools in Germany – The Impact of Social Origin and Gender



Michael Bayer^{a,*}, Sabine Zinn^{b,c}, Christin Rüdiger^d

^a Leibniz Institute for Educational Trajectories (LifBi), Wilhelmsplatz 3, 96047 Bamberg, Germany

^b German Institute for Economic Research (DIW Berlin), Mohrenstrasse 58, 10117 Berlin, Germany

^c Department of Social Sciences, Humboldt University, Berlin, Germany

^d Institute for Educational Quality Improvement (IQB), Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Keywords:

grading
teacher effects
inequality
social background

ABSTRACT

In the German school system, grades are the essential means of performance feedback and assignment. However, little research has been conducted on the factors that determine grades in addition to competencies, and existing findings are poorly replicated. Using data from the representative IQB Trends in Student Performance 2015 survey, our analysis combined a variety of personal and structural characteristics to examine and replicate grade-determining factors in German. In data analysis, we paid particular attention to the data generation processes. The results are interpreted based on Bourdieu's considerations of habitus and status maintenance and Boudon's primary and secondary effects of social origin. The results tie in with the reproduction-theoretical considerations of Bourdieu and Boudon and illustrate the dependence of grades on students' background characteristics.

Introduction

The quality of school grades and the question of what they signify is a topic that crops up in public discourse in Germany at regular intervals. Debates center on the standard of achievement represented by grades, but issues around how grades are determined and certificates awarded are also raised. The question is ultimately one of the extent to which grades and certificates give valid information about students' individual capabilities. A second debate in education policy concerns itself with educational inequality and especially with the impact of social background attributes on the educational development and development prospects of young people (Becker 2003). This association is also the focus of international research, where in addition to structural characteristics of the educational system (c.f. Van der Werfhorst 2018), social origin is considered as relevant for the educational attainment (Bukodi et al. 2021) and competence development (Kulic et al. 2019). Both issues, *inequality* and *ability*, are not only present in public debate. They are also fundamental categories in educational science studies (c.f. Parker et al. 2016). The decision taken in 2003 by the Standing Conference of Education and Cultural Affairs of the *Länder* in the Federal Republic of Germany (*Kultusministerkonferenz, KMK*) to introduce verifiable educational standards for primary and secondary education (KMK, 2010, 2015) can be seen as a reaction to debates about ability and inequality. In parallel with this development, competency-oriented approaches to teaching and learning have been gaining ground (Ertl 2006; Hartong 2012). This paradigm

shift in the German education system throws up questions relating to both changes in the quality of grading and reductions in inequality that may possibly be taking place as a result. Similar considerations have been made in other countries (for the U.S. see Ravitch 2010; for an overall view of the governance implications of the PISA studies in various countries, see Sellar and Lingard 2014). Beyond competency-based teaching, grades, not tests, are the central instrument for determining ability in schools. As grades are still the main selection criterion determining educational transitions within the German school system and onwards (Ditton 2010 for Germany; Erikson and Rudolphi 2010 for Sweden), they merit investigation. Grades are not only the central criterion for transitions between school levels but also for changes between tracks in secondary education. Up to now, studies using representative data to examine changes in grading practices in a targeted fashion, taking account of school types and the federal structures of the German education system, and using results from standardized student achievement testing programs as a reference point for grading have been rare. The German school system, similar to the British one, is highly stratified. This stratification is accompanied by early tracking, where the division of the students takes place at the age of 10 (for a comparative overview see Van de Werfhorst and Mijs 2010 and for the specifics of the German system see Schindler 2021). In addition, in Germany, sovereignty over schools and the respective education system lies at the level of the federal states, resulting in a complex mix of different school types across the country. Studies replicating research on partial aspects of this complex of issues,

* Corresponding author.

E-mail addresses: michael.bayer@lifbi.de (M. Bayer), szinn@diw.de (S. Zinn), christin.ruediger@iqb.hu-berlin.de (C. Rüdiger).

such as grading disparities between boys and girls, are also remarkably rare despite widespread acknowledgment of their value. The present study accordingly seeks to conceptually replicate (Schmidt 2017) existing findings on the links between grades, abilities and backgrounds and, in so doing, to overcome theoretical and methodological shortcomings of earlier studies of grading. In a first step, research discourses on *inequality* and *ability* are taken up and specific findings from educational sociology and from psychology of teaching and learning are discussed. As a second step, and in preparation for our analyses, we discuss the existing evidence on the influences of individual and contextual factors on educational attainment in schools. In doing so, we draw on the seminal approaches of Raymond Boudon and Pierre Bourdieu. On this basis, a multilevel analysis model is developed to investigate factors explaining grades awarded in the subject of German in school reports issued to ninth-grade students. Linguistic competencies have been shown to be strong predictors of further educational and life course outcomes (Duncan et al. 2007) However, in our theoretical framework, they are primarily taken as a key indicator of cultural capital and cultural fit between students and schools (similarly in Smala et al. 2013). To ensure valid statistical inference, we consider both the sampling design of the data used and the nonresponse processes that have led to the final data set in designing and fitting this model. As we reanalyzed the methodology of existing studies comprehensively during the development of our own model, we will discuss how we proceeded in some detail here and link this discussion with critical reflections on previous methodological approaches. In a final step, we draw conclusions from our analysis about the impact of teacher effects on the grading process.

Grading between inequalities and abilities

The PISA studies, most recently in 2019 (OECD 2020), have repeatedly demonstrated that educational success at school level in Germany is strongly dependent on social background, and sociological education research in Germany has reacted by focusing special attention on the reproduction of inequality in and by schools (Becker and Lauterbach 2010). The structure of the German school system with its differentiated school types already represents an important *institutional reproduction mechanism*. A large proportion of the influence of social origin is realized during the transition to secondary schools (*secondary origin effects*) when the school population is distributed (in contrast to countries with a uniform school system) across a range of school types (Becker et al. 2016). Grades research shows that under this school regime the correlation between achievement and grades varies across school types (Baumert et al. 2003), mainly because teachers use the overall performance of class groups as a reference norm (Dickhäuser et al. 2017). Studies also confirm the influence of further factors with a fairly direct bearing on achievement such as self-discipline (Duckworth and Seligman 2006), ability self-concept (Heyder et al. 2017) and effort (Heyder and Kessels 2016). The fact that student attributes less directly linked to performance such as gender (Voyer and Voyer 2014), social background (Helbig and Morar 2018) and migration background (Bonefeld and Dickhäuser 2018) also show an effect on grades highlights the relevance of these dimensions of inequality.

The studies mentioned here also showed that assessment errors were made by teachers in some cases and that these were correlated with the social background and gender of students. Using performance in standardized assessments as a control, Bonefeld et al. (2017) found that children with a migration background face a stable grading disadvantage. Controlling for competencies and motivational-volitional student attributes, Helbig and Morar (2018) found both that children from lower social strata were at a disadvantage that was independent of their level of competency and that the influence of teacher judgments based on teachers' perceptions of the talent and work behavior of children was considerable (*tertiary origin effects*).

Research in the psychology of teaching and learning with a focus on unequal abilities is generally most strongly oriented toward what happens within the school and classroom contexts, and it tends to foreground the diagnostic quality of teachers' judgments of student attributes that are relevant for learning and achievement (Südkamp et al. 2012). It has been established that teachers judge different attributes of students with varying levels of accuracy (Kaiser et al. 2017; Bonefeld et al. 2020, Timmermans et al. 2015). However, the process of evaluating performance in schools is not simply a matter of the accuracy of teacher judgments and the influence of students' social backgrounds on the expectations of teachers. Assessment of student performance in schools takes place for many reasons, as has been highlighted by Gomolla (2012, 26). These multifaceted purposes make unpacking the complexity of grading processes challenging, but they also offer signposts that our analysis and interpretation will follow. They range from the selective function of schools (Fend 1980) to the expression and preservation of teachers' professional ethos that is transported in their autonomous appraisal of students' academic abilities. Without wishing to question the professionalism of teachers, further investigation of grading practices seems to be a desideratum against the background of the striking correlations between educational trajectories and student attributes such as gender, social origin and migration background that have been shown to exist and the objective of reducing inequalities. Established theoretical approaches toward analyzing social origin effects that have yet to be applied to analyzing the determination of grades lend themselves to this endeavor.

Social background, educational achievement at school, and teacher effects

Following the distinction between primary and secondary origin effects established by Boudon (1974) and extended to include tertiary origin effects by Helbig and Morar (2018), sociological education research has been able to show that the effects of social background on educational attainment result from both specific mechanisms and more subtle influences that can be just as significant. While primary origin effects manifest as differences in student competency that can be attributed to family background effects rather than to school effects, secondary origin effects are especially apparent in the German education system in the tracking choices made during the transition from primary school to second level. Tertiary effects are more likely to have an impact on grading processes. They include psychosocial aspects such as the conscientiousness or effort levels of students, but also possible social origin-specific differences in how students are perceived by teachers. In their examination of tertiary origin effects, Helbig and Morar (2018) take up ideas initially put forward by Gresch (2012) and Esser (2016). In Bourdieu's (1982) theory of cultural reproduction, origin effects are not separated into groups in this way: Bourdieu sees both ambitious educational aspirations based on cost-benefit calculations and the form of thinking about educational pathways (in terms of costs and benefits, or otherwise) as an expression of a class-specific habitus (Kramer 2011, 119; Münch 2018, 105–106). Seen from a Bourdieusian perspective, tertiary origin effects reflect the habitual proximity or distance of the students from the school field and ultimately also from the habitus of teachers (Bremer and Lange-Vester 2014). Bourdieu emphasizes that gender is a fundamental dimension of habitus (Bourdieu et al. 1997, 222) and points out that this is reflected in the increased access of girls to upper-track schools and higher education (Bourdieu 2012, 156). From this perspective, social background is an expression of class origin and gender and contributes these factors to the specific form taken by an individual's habitus. Both sources exert substantial influence at levels extending all the way down to the level of classroom interaction (Budde 2014).

The influence of educational background and gender

Taking up analysis by Reckwitz (2017), it can be noted that membership of specific social classes in societies like that of the Federal Republic of Germany is defined, above all, via education. The academic middle class Reckwitz describes as the educated class has gained its position in social space primarily through the education system and expresses strong interest in placing its offspring in advantageous positions within this system and fostering its development. This explains the Bourdieusian conception of the school system as one of the most effective factors perpetuating the existing social order (Bourdieu 2018, 7). From the perspective of this reproduction theory, the performance of students whose parents belong to the educated class tends to be evaluated more positively because the education-focused habitus of this class fits in well with the habitus expected of students by teachers. Two contrasting pathways are mainly responsible for this effect: the class-specific proximity of the academic middle class to the education system is expressed in favorable motivational-volitional (and thus habitual) attributes such as positive attitudes toward learning, and the beliefs and stereotypes held by teachers can lead them to incorporate social background attributes (and gender) into their appraisal of students' performance. Social and gender stereotypes can be seen as a component of perception in the typical habitus of teachers (Lange-Vester and Vester 2018) that not only results in social categorization and in expectations linked to specific categories (Martiny and Fröhlich 2020) but also creates social proximity or distance. Research on stereotypes held by teachers (Pit-Ten Cate and Glock 2018; Tobisch and Dresel 2020) does not usually explore the origins of these stereotypes and cannot clarify the social conditions in which they have arisen. For some time now education research has been discussing findings that show a constant underachievement of boys in terms of performance, grades and certificates (Duckworth and Seligman 2006; Heyder et al. 2017; Voyer and Voyer 2014). Helbig (2012) sees the superior academic performance of girls as the unsurprising result of gender-stereotypical behaviors seen as desirable in schools and peer cultures that foster these behaviors, and others have echoed this: Steinmayr and Spinath (2008) identify high levels of agreeableness and low work avoidance shown by girls as central mediators of grades.

Teacher effects as an expression of perceived proximity/distance to the education system

Building on Boudon with the concept of tertiary origin effects (or, alternatively, on the Bourdieusian conception of habitual proximity) shifts teachers more strongly into the focus of education research by throwing up the question of the mechanisms underlying the reproduction of inequality. Teachers act within a field characterized by education structures, curricular requirements, and education policy. They act as members of a profession, but also as autonomous individuals. The education system possesses the privilege, as Bourdieu and Passeron (1971) pointed out, of training its own potential recruits. Rather unsurprisingly, this induces the effect that teachers, especially, have considerable faith in the fairness and impartiality of their own institution, since they are, as Bourdieu and Passeron (1971, 168) commented, either products of the education system themselves or likely to associate it with opportunities for social advancement. Empirical findings indicate that attention in educational institutions is focused on two groups, in particular: students at the *Gymnasium* school type and girls. One indicator for this is the consistently high proportion of women taking state teaching examinations at the conclusion of their university studies. For the past 20 years, women have made up 70–75 per cent of the candidates taking these examinations. In this respect, European and North American countries are similar, as e.g., Neugebauer et al. (2011) showed. Teachers as a professional group are thus recruited chiefly from the group of women whose prospects are advanced in a special way by ed-

ucational institutions.¹ Starting from these theoretical reflections, we hypothesize that gender-specific and background-specific influences on grading can be explained neither by the structural differences between individual education systems (such as differentiated school types) nor by the composition of specific classes in terms of gender, social background or migration background. At the same time, we acknowledge that both of these groups of factors have been shown to have significant effects on competency scores in standardized tests (Scharenberg 2014). The effect of teacher expectations on students' academic performance has also been widely studied internationally (Wang et al. 2018 for a systematic review). However, most studies focus exclusively on this effect without embedding it in a more general reproduction perspective of educational inequality. In terms of tracking recommendations, Boone et al. (2018) showed that for the transition from primary to secondary schools in the Netherlands the ability context in the class significantly impacts on the respective teacher recommendations. Students in classes with low average abilities were more likely to receive recommendations for academic tracks, even when they had low abilities compared to the overall student population. We assume that gender-specific and background-specific influences can more readily be explained via a recruitment logic inherent to the education system based on social and habitual proximity (Bremer and Lange-Vester 2014).

The subject of German is particularly suited for analysis of this question, since it can be hypothesized that the social reproduction mechanism we have described seems likely to be most pronounced in this subject if we accept Bourdieu's view of language as "the most important part of the cultural heritage. [...] Apart from a lexis and a syntax, each individual inherits from his milieu a certain attitude toward words and their use." (Bourdieu 2018, 27; trans. J.C. Whitehouse, cited from Bourdieu 1974, 40). Following Bourdieu's argument based on the logic of institutions, it seems plausible that the social proximity of specific student groups to teacher expectations rests on milieu-specific similarities rather than secondary socialization processes during teacher training. As such, we do not expect to find differences between the grading practice of teachers who have accessed the profession by traditional routes and career-changers who have taken lateral-entry routes into the profession. Different school types confront members of the profession with varying curricular and learning development requirements. At the same time, the school system operates with a uniform grading scale. In standardized test programs across different school types, this leads to expectations that the differences in academic achievement measured should match the respective school types. As teachers at lower-track schools exploit the power of grades to motivate students just as their colleagues teaching in the upper-track *Gymnasium* do, it can be assumed that the full range of grades is normally used and we can expect differences in grades to be lower than the differences in competency scores found in comparisons across school types.

Making comparisons in Germany's federally-organized education system is not easy, as large differences between *Länder* exist in relation to both the differentiated structures of school systems and the types of school leaving certificates offered in different school types (Helbig and Nicolai 2015). Helbig and Nicolai have proposed a simplified typology that divides the federal states into a total of four groups on the basis of factors such as the possibilities for achieving the *Abitur* university-entrance qualification without attending the *Gymnasium* school type. Three of these school types (the modernized type, the modernized mixed type, and the traditional mixed type) can still be found empirically.

¹ We do not, however, consider that the expression of this effect linked to the habitus of a professional group depends on teacher-student gender matches or mismatches in specific classroom interactions. These arguments are rejected by Neugebauer et al. (2011). They show that feminization thesis little credence as an explanation for boys' poorer grades.

Central research questions and hypotheses

Taking these reported findings and theoretical reflections as our departure point, we will now focus the remainder of our investigation on grading in the subject of German. Drawing on Bourdieu's model of status attainment, we aim to investigate whether our findings fit into a theoretically meaningful interpretative framework and to determine whether earlier results on grading disadvantages affecting children from families with low social status and boys can be replicated. With this objective in mind, we pose and will address the following research question: *How can the determination of grades in the subject of German be explained with reference to the specific structure of the German education system and special consideration of attributes of both students and teachers that research has deemed pertinent?* The research gap we perceive in this context relates to the limitations mentioned above of many prior studies in terms of their scope and data coverage as well as to the question of whether prior results can be replicated. We see a further challenge in re-connecting analyses in this field with theories of society such as Bourdieu's theory of the reproduction of social classes. Both Kronig's (2007) study on the systematic randomness of educational success and Hofstetter's (2017) study on transition recommendations were able to demonstrate clearly that neglecting the social reproduction theory perspective guided by Bourdieu's work has led to research (and especially quantitative analysis) losing sight of the status assignment function performed by schools and the education system as a whole. Starting from this deliberately open research question, we formulate our hypotheses along the structure elaborated in the literature section. Overall, we identified three potential influencing factors: First, the gender of the student, second, his/her academic vs. non-academic educational background, and third, class-related context effects.

H1: Building on existing research results on the influence of gender and on the theoretical aspects considered, we expect to find that girls receive considerably better grades.

H2: Over and above this gender effect, we expect that backgrounds in the academic middle class, operationalized using the criterion of whether students' mothers hold a university degree, will also show a positive effect on grades.

H3: Referring to the findings regarding the referential nature of ability-based class composition for teachers' perceptions of performance, we hypothesize that students in high-performing classes receive comparatively worse grades than those in low-performing classes.

H4: We do not expect individual teacher attributes such as professional experience, gender or age to influence grading, as we see factors specific to teachers as a professional group as more likely to be relevant predictors of how students are perceived and assessed.

H5: In marked contrast to some prior findings, we do not expect to observe an interaction effect between teacher gender and student gender; we do not view the superior grades awarded to girls as a consequence of the proportion of biological women in the teacher population.

The data basis and a description of sampling

The following analyses were computed using data from the IQB Trends in Student Achievement 2015 (IQB-Bildungstrend, Stanat et al. 2018) produced by the Institute for Educational Quality Improvement (IQB). The IQB Trends in Student Achievement 2015 study was conducted in the spring of 2015 with ninth-grade students in all German federal states and tested student competency in German, English, and—in six states—French. The school sampling procedure was stratified by school type and proportional to the size of the ninth grade in each school. In addition, the sampling procedure took account of the fact that the competency levels of students in each general education school type can vary considerably from one federal state to another. To keep the sampling uncertainty in the estimation of competency levels constant across all *Länder*, disproportionately more schools were sampled in those states (such as Berlin and Brandenburg)

where it is known that comparatively large competency differences exist across different school types. In each school drawn, a ninth-grade class was randomly selected for participation. Further details on the sampling procedure are reported in Schipolowski et al. (2016). Weights for the total sample are included in the IQB Trends data. The total sample encompasses $N = 36,542$ students from $N = 1,442$ schools and $N = 1,575$ teachers of German. Of these, $N = 31,594$ students were attending a general education school type. As we assume that the mechanisms our investigation focuses on are varying at different school types, students at special education schools are excluded from our analysis. For $N = 21,813$ students, usable information on the teacher in the subject of German was available (for a total of $N = 978$ teachers in the same number of schools).

Instruments

The IQB Trends in Student Achievement 2015 data captures many attributes on the background of the students, parents, teachers, and schools (scales manual: Schipolowski et al. 2018). The present study was thus able to use constructs at multiple explanatory levels: individual attributes of students and teachers as well as attributes of class groups resp. schools. A comprehensive overview of all variables used, including their medians, standard deviations and proportions (and the proportions of missing values), is given in the online supplement accompanying this article in Table S1.

Focus variables

Individual attributes of students

The grade awarded in the subject of German in student half-year reports for the ninth grade in the school year 2014/2015 is the dependent variable in our modeling. It was recorded, where available, for every student by their teacher.² In Germany the grading system is standardized across most officially recognized school types. There are very few exceptions like Waldorfschulen, but these are not part of the analysis. The gender of students was incorporated into the analysis as a binary variable. The academic background of students' mothers was recorded in the parent questionnaire. (Many studies have shown that the level of education reached by mothers is of central importance for child development, including Stevenson and Baker (1987) and Youniss and Smollar (1985); that is why we focused on this fact.) The variable was trichotomized (a higher university degree, a bachelor's degree or degree from a University of Applied Sciences, a qualification at a lower level). As a further control variable, we used migration background (yes/no), measured based on the generational status of students.

Class composition attributes

As main class composition attribute we focused on the median competency level of the class and the proportion of children with migration backgrounds in the class was aggregated from the individual attributes of the students (after imputation of missing values, see online supplement S2 for details) and was also available as a metric variable. The proportion of girls in the class was aggregated and incorporated into the analyses as a metric variable.

Individual attributes of teachers

As individual-level teacher variables we used gender, professional experience in years and lateral entry. Teacher gender as a dichotomous variable in the analyses. The variable lateral entry (yes/no) was used to record whether teachers had completed undergraduate degree programs

² The teachers reported the grades using the system that was in use at their school. The system in use ("grades," "points," "other" or "none") was also recorded. This made it possible to recode the information given by teachers as grades between 1 and 6 where necessary (with 1 being the best grade and 6 the poorest).

incorporating teacher training and followed by a probation period and a second state examination. The *professional experience* of teachers was recorded in years of experience and thus incorporated into the analyses as a metric variable. All information on teachers was taken from the teacher questionnaire.

Control and design variables

Individual attributes of students

The *competency score in German orthography* is one important control variables in our analysis.³ Orthography is a competency that is primarily taught at school and less subject to the influence of students' family backgrounds than, say, reading competency. In addition, spelling and grammar are components of the subject of German that are covered in all school types in the ninth grade. Orthography is also evaluated in a more standardized fashion than, say, reading competency. The WLE reliability for the competency domain of orthography was 0.895. Details on the educational standards, the competency model used, and the test items are given in the 2015 IQB trends in students achievement report (Becker-Mrotzek et al., 2016).

To measure the *effort level* of students, we constructed a collective (latent) factor from four items of a scale developed at IQB drawing on Litman (2008). The following items (each with a five-point scale) were used: "I often spend hours on a problem and am not satisfied until I have solved it", "When I start things, I finish them, too.", "When I cannot solve a problem, I try even harder." and "I often work zealously on problems that I believe need to be solved." A principal component analysis was performed to construct the factor (see Fahrmeir et al., 2015). We replaced missing values in the 2015 IQB trends report data using multiple imputation. The procedure used for this is described in the online supplement S2. The principal component analysis was performed for each imputed dataset. This allows us to represent the uncertainty resulting from missing item values. The function *fa* from the R package *psych* (package version 1.8.12) was used to derive the factor.

As school achievement can be influenced by abilities that are not specific to certain subjects, the *general cognitive abilities* of students were also incorporated into the analysis, also in the form of a WLE estimator. The instrument used to determine these general cognitive abilities is part of a test on logical thinking (figural aspects) that in turn forms part of the Berlin Test for Measuring Fluid and Crystallized Intelligence in Grades 8–10 (*Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe*, BEFKI; Wilhelm et al. 2014).

Class composition attributes

The IQB Trends in Student Achievement 2015 data contains attributes of students in sampled classes and information on students from parallel classes with diagnosed special educational needs. The IQB weighting does not make it possible to compensate for this kind of oversampling (see Schipolowski et al. 2016). To avoid distorted statistical inference, we used a model-based approach (Winship and Radbill 1994) to compensate for any bias introduced by oversampling. We incorporated the information *student with special needs* and *proportion of students with special needs at school* into the analysis model as control variables.

Furthermore, we controlled for *class size*. The variable was derived as the sum of survey weights of all students in a class and incorporated into the analyses as a metric variable.

School structure attributes

To control for school structure a variable was operationalized using *federal states* as a dichotomous variable in accordance with the

³ In addition to testing German orthography, the IQB Trends in Student Achievement 2015 gathered data on competency in the domains of reading and listening within the subject of German. To avoid problems with high multicollinearity, we have opted to include only the domain of orthography in our analysis.

education system typology of Helbig and Nikolai (2015). This means that the states of Berlin, Brandenburg, Hamburg, Hessen, Mecklenburg-Vorpommern, North Rhine-Westphalia, Rhineland-Palatinate, the Saarland and Schleswig-Holstein are categorized as having modernized (Category 1) school structures in the school year 2009/2010. The states of Lower Saxony, Saxony-Anhalt, Thuringia, Baden-Württemberg, Bavaria and Saxony were categorized as having mixed (Category 2) modernized or traditional school structures. The *Gymnasium school type* flowed into the analyses as a binary variable (yes/no).

Method and analysis model

We used a weighted multivariate regression model to analyze factors contributing to the half-year grades assigned in the subject of German. As the half-year grade in the subject of German, the dependent variable in our model, is an ordinal variable, an *ordered logit* regression with represents a suitable statistical modeling framework. The intra-class correlation of the German grade at the teacher level (controlling for all observed teacher characteristics) is 17.6%. This indicates a large amount of unobserved heterogeneity at this level. To avoid bias in the variance estimates, we extend the ordered logit model with a related random effect. (For those interested, the formula for this statistical model can be taken e.g., from the paper of Hedeker and Gibbons (1994) on random effects ordinal regression models for multilevel analysis.) We used the *meologit* function in Stata, Version 17, to fit the analysis model. Before implementing this step, however, missing values in the student data relating to missing student variables (i.e., *item nonresponse* at the student level) were replaced using multiple imputation. For the weighted *ordered logit* analyses, missing teacher information (i.e., *unit nonresponse*) was compensated for using nonresponse weight adjustments. Our decision to compensate for *item nonresponse* with multiple imputation and for *unit nonresponse* with weight adjustments follows the recommendations made by Brick and Kalton (1996) and Brick (2013). Multiple imputation (as a model-based procedure) is well-suited to data situations in which one knows a lot about the entity with the missing values (in this case, the student) and can represent the correlation structure behind the missing values successfully. Weighting, in contrast, draws on considerably less information; as a design-based approach, it is suitable for compensating for entire missing entities. The online supplement S2 describes the approach used to handle missing data in detail.

To ensure that our results can be replicated, the entire data preparation, imputation and weighting syntax in R and the syntax for fitting the models in Stata and combining the results using Rubin's Rules is publicly available and may be downloaded from <https://github.com/bieneSchwarze/AcademicPerformanceAssesmt->

Results

The (pooled) results of our analyses are contained in Table 1.

Gender and educational background

The results of the regression analysis clearly show that the associations between the student's gender and grade are very strong: Being a female student significantly increases the likelihood of receiving grade 1 (which would correspond to A-level in the U.S. and the UK) and decreases the likelihood of receiving grade 5 or 6 (which would correspond to F-level in the U.S. and the Fail classification in the UK). This confirms our first hypothesis H1. That there is a pronounced and persistent gap between boys and girls in terms of linguistic competencies has also recently been shown by van Heck et al. (2019) (with Germany ranking in the middle of OECD countries in this regard). However, their analyses focused exclusively on competencies determined via tests. Thus, in our analyses, we find that there is a (persistent) gender effect that goes beyond competencies. Based on our theoretical considerations, we interpret this effect as a fit effect that arises from the teacher's perception

Table 1

. Estimated (non-standardized) beta-coefficients with 95% confidence intervals of the *ordered logit* model explaining the mark in the subject of German in the half-year report for the ninth grade in spring 2015

Variables	reference	Effects (non-standardized)	95% confidence interval
Level: students			
Competency score in German orthography	–	-0,548*	(-0,656; -0,440)
BEFKI (general cognitive abilities)	–	-0,169*	(-0,210; -0,128)
Effort level (latent factor)	–	-0,205*	(-0,282; -0,127)
Gender	male		
female		-0,937*	(-1,124; -0,749)
Mother with tertiary educational qualification	no		
yes, university degree		-0,186*	(-0,574; -0,144)
yes, college resp. bachelor degree		-0,359	(-0,409; 0,037)
Student has migration background	no		
yes		0,264*	(0,142; 0,386)
Interaction: competency and mother with tertiary educational qualification	no		
yes, university degree		-0,099	(-0,293; 0,095)
yes, college resp. bachelor degree		-0,039	(-0,224; 0,146)
Interaction: female and mother with tertiary educational qualification	no		
yes, university degree		-0,117	(-0,421; 0,187)
yes, college resp. bachelor degree		-0,008	(-0,326; 0,311)
Design: student has special educational needs ^(a)	no		
Yes		0,713*	(0,258; 1,168)
Level: school resp. class			
Median competency level in German orthography of class		0,283*	(0,013; 0,553)
Class size	–	0,009	(-0,014; 0,032)
Proportion of students with migration background in class	–	0,384	(-0,145; 0,912)
Proportion of female students in class	–	0,103	(-0,421; 0,626)
Class in <i>Gymnasium</i> school type	no		
yes		-0,333*	(-0,601; -0,066)
School structure	mixed modernized / mixed traditional		
modernized		0,169	(-0,069; 0,282)
Design: Number of students with special education needs in school ^(a)	–	-0,003	(-0,011; 0,004)
Level: Teacher			
Professional experience (years)	–	-0,001	(-0,009; 0,007)
Teacher gender	male		
female		0,029	(-0,177; 0,235)
Lateral entry	yes		
no		-0,191	(-0,452; 0,071)
Level: students and teacher			
Interaction: Teacher is female and student is female	no		
yes		-0,086	(-0,297; 0,124)
Intra-class correlation on teacher level ^(b)	17,6% (SD 0,01)		
Sample size (unweighted / weighted)			
Students	21.813 / 790.035		
Teachers resp. class or schools	978 / 35.490		
Pseudo R ² ^(c)	0,34 (SD 0,01)		

Notes: * significant with $p < 0.05$. (a) Variable only for purpose of recognition of the sampling design. The effect could not be interpreted meaningfully, because the assessment process of students with special educational needs is different from other students. (b) Arithmetic means out of 20 analyses with imputed datasets (standard deviation in brackets). (c) Pseudo R-Square from McKelvey and Zavoina (1975) further developed by Langer (2017) for multilevel models. Arithmetic means out of 20 analyses with imputed datasets (standard deviation in brackets).

of gender-specific behaviors. That is, we do not interpret this effect as a simple stereotype effect expressing the teacher's belief that girls have better language skills than boys.

Following Bourdieu's reproduction theory, we expect students' educational background to have an effect on their grades, independent of their competence score and cognitive ability (hypothesis 2). In the analyses, we find that, controlling for all other variables, the probability of receiving a very good grade increases if the student stems from a home in which the mother has a university degree. This corroborates our second hypothesis. However, considering the interaction effect of student gender and mother's educational background we see that at the same time these effects are independent from each other. In other words, the impact of a student's gender on his/her grade does not change with mother's educational attainment.

From our control variables, we see that students with a migration background are significantly disadvantaged with respect to grades, controlling for competency in orthography, general cognitive abilities and motivational attributes (thus replicating previous similar results, see Diehl et al. 2016; Bonefeld et al. 2017). We also find that students with special educational needs receive significantly lower grades than

their peers, even when orthographic competence and cognitive ability are taken into account. However, the latter result should be interpreted with caution, as the assessment procedure for students with special educational needs differs from that for other students.

Class and school structure and composition attributes

Based on previous findings on the effects of the class context on teacher evaluation, we hypothesized that grades tend to be worse in classes with higher average achievement or ability levels than in classes with lower average achievement levels (hypothesis H3). The findings of our regression analysis confirm this assumption. Thus, controlling for the other characteristics, the probability of receiving the top grade decreases when the student is in a high-performing class (measured as the median of orthographic skills in the class).

The other compositional characteristics of the class that we consider in our model (i.e., proportion of female students and proportion of students with a migration background) do not show any significant effects on grades. However, we find a school form effect that we did not expect in this form. Students who are in a *Gymnasium* class get significantly

better grades than students in other types of school, controlling for all other characteristics. This finding is contrary to expectation: Due to the grading scale across school types on the one hand and the lower curricular requirements in the non-Gymnasial school types on the other hand, such an effect should actually not exist. Thus, this effect suggests that there is either some kind of grade inflation in Gymnasiums or that students are systematically graded lower in other school types. In general, this finding should be examined in more detail in further work.

Individual-level teachers attributes

We find that, under the control of student's attribute and class composition characteristics, none of the examined teacher characteristics (i.e., professional experience in years, gender, lateral entry) shows any significant impact on their grading of students. This is also true for the interaction effect of student and teacher gender. This result confirms our hypothesis 4 and is congruent with our expectation that teachers, as members of a profession with its own implicit judgment strategies, foster the talent of students they perceive as prospective replacements. Since gender-differentiated perceptions cannot be explained in terms of the (biological) gender of a teacher, our regression results thus also confirm hypothesis 5.

The fact that professional experience has no influence on grading suggests that the effects we found (beyond the influence of competence) are the result of a relatively stable professional conception. In other words, teachers' profession image seems to be strongly anchored in their mindsets. Apparently, it is still unaffected by the changeover to competency-based standards.

Discussion and limitations

The current study addresses the question to which extent in Germany ninth-grade grades in German (i.e., the subject in which the official national language is taught) can be explained by student characteristics such as gender and social origin, class composition, and teacher characteristics, controlling for students' linguistic competence and the cognitive ability, among others. In addition to drawing on work and ideas formulated by Bourdieu on social reproduction theory, replicating earlier studies drawing on representative data for ninth-grade students in the year 2015 was of central interest.

Two essentials set our study apart from other work in the field. *Firstly*, we were able to replicate some findings from previous research in this area. The picture revealed by our analysis is similar to that shown by other studies in that we found a strong effect of student gender and family educational background on the grade awarded in German. But the large size of the effects we found was striking, and it can be taken that these effect sizes are robust on account of the methodological approach used. Our findings show that grades in German, a subject of major significance for the reproduction of inequality, are dependent on domain-specific competency (in our analyses: in orthography) only to a limited extent and that student (background) attributes less directly linked to performance also contribute to explaining grades. This finding can also be derived theoretically from Bourdieu's thinking on social reproduction. Our study confirms the diagnosis that boys are the new losers in education, not only in terms of competency development and distribution, but also in relation to grading practice.

The fact that motivational-volitional characteristics (such as self-assessed effort) impact on grading is an established finding in international research (Heyder and Kessels 2016). It is also well known that gender differences are partly due to this circumstance (see, e.g., Jackson and Dempster 2009). While Wokman and Heyder (2020) see these gender differences as the result of different social costs, we interpret them as the result of the reproductive urge inherent in education systems in Bourdieu's sense (Bourdieu and Passeron 2007). That is, we conjecture that the German education system gives students from certain milieus (namely, the academic middle class) advantages over

other student groups, as manifested in and through teachers' perceptions and evaluations of them. This theoretical approach is strengthened by Brookhart et al. (2016) who stated in a broad meta-study on grading processes "that grades typically represents a mixture of multiple factors that teachers value" (Brookhart et al. 2016, p. 843). With this reproduction theory perspective on our findings, we link to analyses in which differences between the educational trajectories of students from different social origins have been examined for their causes. While Van de Werfhorst and Hofstede (2007) question the usefulness of the Bourdieu's cultural reproduction approach in explaining educational transitions, Jæger and Breen (2016) show the empirical relevance of Bourdieu's central reproduction assumption (by means of their dynamic model of cultural reproduction). However, in one of their conclusions Jaeger and Breen point out that a more precise conceptualization of the influences of cultural capital is still needed. Therefore, we also see our analyses as a contribution to further elucidating the pathways of reproduction-relevant characteristics.

While a large number of studies see students' family educational background as mediated primarily through cultural and educational practices on academic performance, our analyses, controlling for ability in German, show that family educational background gains influence on teachers' assessment in ways other than an ability-related pathway. In their analyses of a Danish sample of twins, Jæger and Møllegaard (2017) found no direct effect of cultural capital (as a form of operationalizing the educational background of parents) on yearly grades, but likewise on final grades in secondary school. Moreover, our results fit particularly well with their finding that it is especially students from tertiary educated families for whom cultural capital has a direct impact on school success. In this context, it is important to keep in mind that grades are a central predictor of further educational progress not only in Germany (see Atkinson and Geiser 2009 for the United States and Vidal Rodeiro and Zanini 2015 for the UK). Our study thus contributes empirical clarification to the research gap of a missing conceptualization of cultural reproduction with respect to the particularities of a specific educational system. We achieve this by showing that beyond the differences between educational systems in terms of their differentiation and standardization, (in Germany) there are educational system effects that systematically favor certain groups of students. Therefore, offsetting the inequality-causing effects of tracking structures through further standardization, as suggested by Bol et al. (2014), is very unlikely to be successful in our view. The theoretical perspective we propose to interpret the effects found, *secondly*, takes up theoretical thinking and hypotheses on structures and on the teaching profession that can, to some extent, only be made indirectly visible in a model based on individual data, but can link up with interpretations gained from qualitative research. Common attitudes and perceptions shared by members of the teaching workforce can be identified, if only because of the mere fact that teachers have followed very similar educational pathways which lead them to have similar expectations of students across all school types and to hold similar education-related stereotypes (Helsper 2018). While qualitative approaches (Bremer and Lange-Vester 2014) have been able to reconstruct diverse forms of teacher habitus, it can be shown with our data that teachers' preference for certain groups (in the sense that family educational background and gender effects show an advantage when orthographic competency and self-assessed effort are controlled for) can be interpreted as a relationship of proximity or distance to the institutions of the education system.

To take this interpretative approach further, however, it would be necessary to know more about the social background of the teachers. Over and above this, information about other aspects of the education and training of teachers that is relevant for their attitudes and beliefs also represents potential explanatory components that should be considered in investigations applying Bourdieu's reproduction theory approach to teachers.

With a view to further research, some further limitations of the present study should be mentioned. We found in our analysis that a

special educational needs diagnosis has a significant negative effect on receiving the top grade in the school subject “German” (while controlling for competency, fluid intelligence, and so on). However, this finding cannot be interpreted in greater depth due to the different grading logic used to mark the work of students with special educational needs. In addition, only the test in the competency domain of orthography was used to represent achievement in the subject of German. Important aspects of the curricula for teaching German were not considered. It can be assumed, moreover, that the assessments of educational attainment conducted by teachers are also influenced by the reference norm orientations of teachers and by the work behavior (Dickhäuser et al. 2017) and the classroom and homework behavior of students (Brookhart et al. 2016). However, we were unable to consider these factors in our analysis, as corresponding variables were not included in the 2015 IQB Trends data. We believe we have shown that the process of grading remains an important research field, especially with a view to current education policy efforts in connection with the implementation of new competency-oriented educational standards. We note that our study also served as a replication study. Unfortunately, this type of study is still far too neglected, although it is precisely replication studies that serve to reveal the persistence of grievances such as the persistence of educational inequalities.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijedro.2021.100101.

References

- Atkinson, R. C., & Geiser, S. (2009). Reflections on a Century of College Admissions Tests. *Educational Researchers*, 38(9), 665–679.
- Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten—institutionelle Bedingungen des Lehrens und Lernens. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000—Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (pp. 261–331). Opladen: Verlag Leske + Budrich.
- Becker-Mrotzek, M., Böhme, K., Bulut, N., Hunger, S., Jost, J., Mörs, M., & Stanat, P. (2016). Integrierte Kompetenzstufenmodelle im Fach Deutsch. In P. Stanat, K. Böhme, S. Schipolowski, & N. Haag (Eds.), *IQB-Bildungstrend 2015 Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich* (pp. 95–126). Münster, New York: Waxmann.
- Becker, R. (2003). Educational Expansion and Persistent Inequalities of Education. *European Sociological Review*, 19(1), 1–24.
- Becker, R., & Lauterbach, W. (2010). Bildung als Privileg—Ursachen, Mechanismen, Prozesse und Wirkungen. In R. Becker, & W. Lauterbach (Eds.), *Bildung als Privileg. Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (pp. 11–49). Wiesbaden: Springer.
- Becker, M., Neumann, M., & Dumont, H. (2016). Recent Developments in School Tracking Practices in Germany: An Overview and Outlook on Future Trends. *ORBIS SCHOLAE*, 10(3), 9–25.
- Bol, T., Witschge, J., Van de Werfhorst, H. G., & Dronkers, J. (2014). Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries. *Social Forces*, 92(4), 1545–1572.
- Bonefeld, M., & Dickhäuser, O. (2018). (Biased) Grading of Students' Performance: Students' Names, Performance Level, and Implicit Attitudes. *Front Psychol*, 9(481), 1–13.
- Bonefeld, M., Dickhäuser, O., Janke, S., Praetorius, A., & Dresel, M. (2017). Migrationsbedingte Disparitäten in der Notenvergabe nach dem Übergang auf das Gymnasium. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 49(1), 11–23.
- Bonefeld, M., Dickhäuser, O., & Karst, K. (2020). Do preservice teachers' judgments and judgment accuracy depend on students' characteristics? The effect of gender and immigration background. *Social Psychology of Education*, 23, 189–216.
- Boone, S., Thys, S., Van Avermaet, P., & Van Houtte, M. (2018). Class composition as a frame of reference for teachers? The influence of class context on teacher recommendations. *British Educational Research Journal*, 44(2), 274–293.
- Boudon, R. (1974). *Education, opportunity, and social inequality*. New York, NY: Wiley.
- Bourdieu, P. (1974). The school as a conservative force: scholastic and cultural inequalities. In J. Eggleston (Ed.), *Contemporary research in the sociology of education* (pp. 32–46). London: Methuen.
- Bourdieu, P. (1982). *Die feinen Unterschiede [La distinction: Critique sociale du jugement]*. Frankfurt A.M.: Suhrkamp.
- Bourdieu, P. (2012). *Die männliche Herrschaft [La domination masculine]*. Frankfurt a.M.: Suhrkamp.
- Bourdieu, P. (2018). *Bildung. Schriften zur Kultursoziologie 2*. Berlin: Suhrkamp.
- Bourdieu, P., & Passeron, J.-C. (1971). *Die Illusion der Chancengleichheit. Untersuchungen zur Soziologie des Bildungswesens am Beispiel Frankreichs*. Stuttgart.
- Bourdieu, P., & Passeron, J.-C. (2007). *Die Erben. [Les héritiers.]*. Konstanz: UVK.
- Bourdieu, P., Dölling, I., & Steinrück, M. (1997). Eine sanfte Gewalt. In I. Dölling, & B. Kraus (Eds.), *Ein alltägliches Spiel. Geschlechterkonstruktion in der sozialen Praxis* (pp. 218–230). Frankfurt a. M.: Suhrkamp.
- Bremer, H., & Lange-Vester, A. (2014). Die Pluralität der Habitus- und Milieuformen bei Lernenden und Lehrenden. In W. Helsper, R.-T. Kramer, & S. Thiersch (Eds.), *Schülerhabitus* (pp. 56–81). Wiesbaden: Springer VS.
- Brick, J. M. (2013). Unit nonresponse and weighting adjustments: a critical review. *J. Off. Statist.*, 29, 329–353.
- Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical methods in medical research*, 5(3), 215–238.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86(4), 803–848.
- Budde, J. (2014). Das Konzept des männlichen Habitus. Möglichkeiten und Grenzen für die Analyse von Unterrichtspraktiken von Schülern. In W. Helsper, R.-T. Kramer, & S. Thiersch (Eds.), *Schülerhabitus* (pp. 82–98). Wiesbaden: Springer VS.
- Bukodi, E., Goldthorpe, J. H., & Zhao, Y. (2021). Primary and secondary effects of social origins on educational attainment: New findings for England. *British Journal of Sociology*, 72, 627–650.
- Dickhäuser, O., Janke, S., Praetorius, A.-K., & Dresel, M. (2017). The Effects of Teachers' Reference Norm Orientations on Students' Implicit Theories and Academic Self-Concepts. *Zeitschrift für Pädagogische Psychologie*, 31(3-4), 205–219.
- Diehl, C., Hunkler, C., & Kristen, C. (2016). *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*. Wiesbaden: VS Verlag.
- Ditton, H. (2010). Selektion und Exklusion im Bildungssystem. In G. Quenzel, & K. Hurrelmann (Eds.), *Bildungsverlierer* (pp. 53–71). Wiesbaden: VS Verlag.
- Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98, 198–208.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446.
- Erikson, R., & Rudolphi, F. (2010). Change in Social Selection to Upper Secondary School—Primary and Secondary Effects in Sweden. *European Sociological Review*, 26(3), 291–305.
- Ertl, H. (2006). Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany. *Oxford Review of Education*, 32(5), 619–634 Special Issue: Comparative Inquiry and Educational Policy Making.
- Esser, H. (2016). Sorting and (much) more: prior ability, school effects and the impact of ability tracking on educational inequalities in achievement. In A. Hadjar, & C. Gross (Eds.), *Education Systems and Inequalities: International Comparisons* (pp. 95–114). Bristol: Policy Press Scholarship.
- Fahrmeir, L., Hamerle, A., & Tutz, G. (Eds.). (2015). *Multivariate statistische Verfahren*. Berlin: De Gruyter.
- Fend, H. (1980). *Theorie der Schule*. München: Urban & Schwarzenberg.
- Gomolla, M. (2012). Leistungsbeurteilung in der Schule: Zwischen Selektion und Förderung, Gerechtigkeitsanspruch und Diskriminierung. In S. Fürstenau, & M. Gomolla (Eds.), *Migration und schulischer Wandel: Leistungsbeurteilung* (pp. 25–50). Wiesbaden: Springer VS.
- Gresch, C. (2012). *Der Übergang in die Sekundarstufe I*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartong, Sigrig (2012). *Basiskompetenzen statt Bildung? Wie PISA die deutschen Schulen verändert hat*. Frankfurt a. M.: Campus.
- Hedeker, D., & Gibbons, R. D. (1994). A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics*, 50(4), 933–944.
- Helbig, M. (2012). Warum bekommen Jungen schlechtere Schulnoten als Mädchen? Ein sozialpsychologischer Erklärungsansatz. *Zeitschrift für Bildungsforschung*, 2, 41–54.
- Helbig, M., & Morar, T. (2018). Why Teachers Assign Socially Unequal Marks. A Case for Establishing Tertiary Origin Effects in the Model of Primary and Secondary Effects of Educational Research. *International Journal of Learning, Teaching and Educational Research*, 17(7), 1–25.
- Helbig, M., & Nikolai, R. (2015). *Die Unvergleichbaren: Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949*. Bad Heilbrunn: Julius Klinkhardt.
- Helsper, W. (2018). Lehrerhabitus. In A. Paseka, M. Keller-Schneider, & A. Combe (Eds.), *Ungewissheit als Herausforderung für pädagogisches Handeln* (pp. 105–140). Wiesbaden: Springer VS.
- Heyder, A., & Kessels, U. (2016). Boys Don't Work? On the Psychological Benefits of Showing Low Effort in High School. *Sex Roles*, 77(1-2), 72–85.
- Heyder, A., Kessels, U., & Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions? *British Journal of Educational Psychology*, 87(2), 205–223.
- Hofstetter, D. (2017). *Die schulische Selektion als soziale Praxis. Aushandlungen von Bildungsentscheidungen beim Übergang von der Primarschule in die Sekundarstufe I*. Weinheim: Beltz Juventa.
- Jæger, M. M., & Breen, R. (2016). A Dynamic Model of Cultural Reproduction. *American Journal of Sociology*, 121(4), 1079–1115.
- Jæger, M. M., & Mollegaard, S. (2017). Cultural capital, teacher bias, and educational success: New evidence from monozygotic twins. *Social Science Research*, 65, 130–144.
- Jackson, C., & Dempster, S. (2009). 'I sat back on my computer . . . with a bottle of whisky next to me': constructing 'cool' masculinity through 'effortless' achievement in secondary and higher education. *Journal of Gender Studies*, 18, 341–356.
- Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. *Journal of Educational Psychology*, 109(6), 871–888.

- Kramer, R.-T. (2011). *Abschied von Bourdieu?*. Wiesbaden: Springer VS.
- Kronig, W. (2007). *Die systematische Zufälligkeit des Bildungserfolgs. Theoretische Erklärungen und empirische Untersuchungen zur Lernentwicklung und zur Leistungsbewertung in unterschiedlichen Schulklassen*. Bern: Haupt Verlag.
- Kulic, N., Skopek, J., Treiventi, M., & Blossfeld, H.-P. (2019). Social Background and Children's Cognitive Skills: The Role of Early Childhood Education and Care in a Cross-National Perspective. *Annual Review of Sociology*, 45(1), 557–579.
- Kultusministerkonferenz. (2010). *Konzeption der Kultusministerkonferenz zur Nutzung der Bildungsstandards für die Unterrichtsentwicklung*. Cologne: Wolters Kluwer.
- Kultusministerkonferenz. (2015). *Gesamtstrategie Bildungsmonitoring der Kultusministerkonferenz zum Bildungsmonitoring*. Cologne: Wolters Kluwer.
- Lange-Vester, A., Vester, M., et al. (2018). Lehrpersonen, Habitus und soziale Ungleichheit in schulischen Bildungsprozessen. In K.-H. Braun, et al. (Eds.), *Erziehungswissenschaftliche Reflexion und pädagogisch-politisches Engagement* (pp. 159–183). Wiesbaden: Springer VS.
- Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, 44, 1585–1595.
- Martiny, S. E., & Fröhlich, L. (2020). Ein theoretischer und empirischer Überblick über die Entwicklung von Stereotypen und ihre Konsequenzen im Schulkontext. In S. Glock, & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 1–32). Wiesbaden: Springer VS.
- Münch, R. (2018). *Der bildungsindustrielle Komplex*. Frankfurt a.M.: Beltz Juventa.
- Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the Myth of the Same-Sex Teacher Advantage. *European Sociological Review*, 27(5), 669–689.
- OECD. (2020). Students' socio-economic status and performance. *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing.
- Parker, P. D., Jerrim, J., Schoon, I., & Marsh, H. W. (2016). A Multination Study of Socioeconomic Inequality in Expectations for Progression to Higher Education: The Role of Between-School Tracking and Ability Stratification. *American Educational Research Journal*, 53(1), 6–32.
- Pit-Ten Cate, I. M., & Glock, S. (2018). Teacher's attitudes toward students with high- and low-educated parents. *Soc Psychol Educ*, 21, 725–742.
- Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. New York: Basic Books.
- Reckwitz, A. (2017). *Die Gesellschaft der Singularitäten*. Berlin: Suhrkamp.
- Scharenberg, K. (2014). Schule und Schulklasse als soziale Kontexte der Entwicklung im Jugendalter. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(S1), 317–348.
- Schindler, S. (2021). Educational differentiation in secondary education and labour-market outcomes. *Longitudinal and Life Course Studies*, 12(3), 271–277.
- Schipolowski, S., Haag, N., Böhme, K., & Sachse, K. (2016). Anlage, Durchführung und Auswertung des IQB-Bildungstrends 2015. In P. Stanat, K. Böhme, S. Schipolowski, & N. Haag (Eds.), *IQB-Bildungstrend 2015 Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich* (pp. 95–126). Münster, New York: Waxmann.
- Schipolowski, S., Haag, N., Milles, F., Pietz, S., & Stanat, P. (2018). *IQB-Bildungstrend 2015. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente in den Fächern Deutsch und Englisch*. Berlin: Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen.
- Schmidt, S. (2017). Replication. In M. C. Makel, & J. A. Plucker (Eds.), *Toward a More Perfect Psychology: Improving Trust, Accuracy, and Transparency in Research* (pp. 233–253). Washington, District of Columbia: American Psychological Association.
- Sellar, S., & Lingard, B. (2014). The OECD and the expansion of PISA: new global modes of governance in education. *British Educational Research Journal*, 40(6), 917–936.
- Smala, S., Paz, J. B., & Lingard, B. (2013). Languages, cultural capital and school choice: distinction and second-language immersion programmes. *British Journal of Sociology of Education*, 34(3), 373–391.
- Stanat, P., Böhme, K., Schipolowski, S., Haag, N., Weirich, S., Sachse, K., Hoffmann, L., & Federlein, F. (2018). *IQB-Bildungstrend Sprachen 2015 (IQB-BT 2015) Version: 2. IQB—Institut zur Qualitätsentwicklung im Bildungswesen Dataset*.
- Steinmayr, R., & Spinath, B. (2008). Sex Differences in School Achievement: What Are the Roles of Personality and Achievement Motivation? *European Journal of Personality*, 22(3), 185–209.
- Stevenson, D. L., & Baker, D. P. (1987). The Family-School Relation and the Child's School Performance. *Child Development*, 58(5), 1348–1357.
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *Br J Educ Psychol*, 85(4), 459–478.
- Tobisch, A., & Dresel, M. (2020). Fleißig oder faul? Welche Einstellungen und Stereotype haben angehende Lehrkräfte gegenüber Schüler*innen aus unterschiedlichen sozialen Schichten?. In S. Glock, & H. Kleen (Eds.), *Stereotype in der Schule* (pp. 133–158). Wiesbaden: Springer VS.
- Van de Werhorst, H. G. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. *Research in Social Stratification and Mobility*, 58, 22–32.
- Van de Werhorst, H. G., & Hofstede, S. (2007). Cultural capital or relative risk aversion? Two mechanisms for educational inequality compared. *The British Journal of Sociology*, 58(3), 391–415.
- Van de Werhorst, H. G., & Mijs, J. J. B. (2010). Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective. *Annual Review of Sociology*, 36(1), 407–428.
- Van Hek, M., Buchmann, C., & Kraaykamp, G. (2019). Educational Systems and Gender Differences in Reading: A Comparative Multilevel Analysis. *European Sociological Review*, 35(2), 169–186.
- Vidal Rodeiro, C. L., & Zanini, N. (2015). The role of the A* grade at A level as a predictor of university performance in the United Kingdom. *Oxford Review of Education*, 41(5), 647–670.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol Bull*, 140(4), 1174–1204.
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3-5), 124–179.
- Wilhelm, O., Schroeders, U., & Schipolowski, S. (2014). *Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 8. bis 10. Jahrgangsstufe (BEFKI 8-10)*. Göttingen: Hogrefe.
- Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research*, 23(2), 230–257.
- Wokman, J., & Heyder, A. (2020). Gender achievement gaps: the role of social costs to trying hard in high school. *Social Psychology of Education*, 23, 1407–1427.
- Youniss, J., & Smollar, J. (1985). *Adolescent relations with mothers, fathers, and friends*. Chicago, Ill.: University of Chicago Press.