Studies in Educational Evaluation 86 (2025) 101495

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

# Less alarming than assumed: New insights on diversion effects in German secondary education

Felix Bittmann [1]

*Leibniz Institute for Educational Trajectories, Wilhelmsplatz 4, Bamberg 96047, Germany*

ARTICLE INFO

*Keywords:*
German secondary education
Tracking
Diversion
Learning progress
Social stratification
Statistical balancing

ABSTRACT

Reforms have aimed to make the German educational system more permeable. Today, obtaining the highest level of secondary education is possible regardless of the track to which a student has transferred. However promising, research has indicated that these hopes of easier upward mobility have only been partially fulfilled, as many students who began their educational journeys on a non-academic track displayed diversion effects—that is, a loss of initially high aspirations. Panel data allows for testing whether educational outcomes are influenced by the choice of school track among highly motivated students. Our results show that competencies and satisfaction are not affected once self-selection into tracks is accounted for. Regarding the attainment of the final degree, the difference between the academic track (Gymnasium) and the non-academic tracks amounts to approximately 13 percentage points. These findings suggest that diversion effects are smaller than previously reported. Policy-makers, in particular, should be aware of these differences.

## 1. Introduction

Over the past decades, the German institutional educational landscape has undergone significant transformation (Dudek, 1993). In the past, initial track placement in secondary education was highly indicative of final educational attainment. However, a series of educational reforms has aimed to make the system more open and permeable. By introducing new (bridging) school types—particularly in upper secondary education—and facilitating easier transitions between tracks to support sequential educational upgrading, the traditionally lower tracks have shed their former dead-end character. As a result, initial track placement after primary school has become less definitive (Schindler, 2014). While this development is widely seen as a positive step toward reducing social inequality, recent research raises doubts about the success of these reforms. A key criticism is the potential for diversion effects, which may even exacerbate social inequality (Schindler & Bittmann, 2021). This refers to a pattern in which bright but risk-averse students—especially those from socially disadvantaged backgrounds—opt not to pursue the direct academic path (Gymnasium) but instead follow a sequential upgrading route perceived as less risky. However, even when these students begin secondary education with high aspirations, they may lose motivation over time within non-academic tracks and ultimately attain a lower qualification than they initially intended. Some studies have attempted to assess these diversion effects using intermediate outcomes, such as aspirations or academic performance (Bittmann & Schindler, 2021; Holtmann et al., 2024; Traini et al., 2021), but definitive findings based on final educational qualifications have been lacking. We address this gap by analyzing actual educational outcomes using recently released, long-term panel data. This allows us to examine whether students who were highly motivated at the beginning of secondary school to obtain the highest qualification (the Abitur) were ultimately able to achieve their goal. This constitutes our main research question: a direct, empirical assessment of the severity of diversion effects in German secondary education. Using large-scale panel data and following students over a ten-year period, we are able to generate meaningful estimates. To approximate the effects of school track placement, we apply a matching approach to control for (self-)selection into tracks, ensuring comparability between groups at the outset of secondary education. To present a comprehensive picture, we examine both the main outcome—whether the Abitur was obtained—and key antecedents, such as academic achievement and satisfaction with school and life. We also conduct an extensive series of robustness checks to confirm the stability of our findings and to provide readers with a clear sense of the statistical uncertainty surrounding our conclusions.

*E-mail address:* felix.bittmann@lifbi.de.
[1] ORCID: 0000–0003-0802–5854

https://doi.org/10.1016/j.stueduc.2025.101495
Received 10 April 2024; Received in revised form 11 July 2025; Accepted 22 July 2025
Available online 26 July 2025

## 2. Theoretical considerations

### 2.1. The German educational system

The traditional German educational system features an early and rigid between-school tracking structure (Eckhardt, 2017). Students are assigned to one of up to four different tracks directly after primary school—which typically lasts four years in most federal states—when they are around ten years old. The system is designed to sort students based on academic ability and is partially informed by a teacher's recommendation given in the final year of primary education. This (subjective) recommendation is based on prior academic performance and grades. In most federal states, parents are free to follow or disregard the recommendation, although in some states it remains binding and can only be overruled in exceptional cases. Secondary education typically offers three or four school types, depending on the federal state. Traditionally, the *Hauptschule* (lower secondary school) was designed to prepare students for blue-collar occupations and generally lasts five additional years. In recent decades, however, this track has become increasingly marginalized and carries a stigma, often perceived as serving students from lower social backgrounds with weaker academic performance. Consequently, most federal states have abolished, merged, or re-modeled this track (Hurrelmann, 2013). The second option, the intermediate secondary school (*Realschule*), is intended to prepare students for white-collar and non-tertiary occupations, traditionally linked to working in banks, offices, or the service sector. Nowadays, many federal states have also introduced comprehensive schools (*Gesamtschulen*) to combine multiple tracks within one building and facilitate easier switching between tracks. Some forms of these types combine multiple tracks within one building or compound (cooperative form). In contrast, the second form completely abandons any tracking and educates all students in integrated classrooms (integrative form). Finally, the academic track (*Gymnasium*) has a century-old tradition and has usually served as the elite schooling form, preparing students for academic professions. While the share of students in this track has been traditionally low, nowadays, around 50 % of the student cohort enter this track, rendering it somewhat the "default" as it is also present in all federal states.[1] The academic track is the only pathway that directly awards higher education eligibility (HEE; Abitur) after eight or nine years of schooling, thereby granting access to all tiers of tertiary education. Students in this track who do not drop out due to insufficient academic performance and who pass the final examinations are automatically awarded the Abitur, with no additional decisions required. This marks a major distinction from all other school tracks, which typically lead only to lower secondary qualifications. The only exception is comprehensive schools, which include internal academic tracks; in this study, we classify these as a variant of the academic track. In recent decades, however, broad educational reforms have aimed to eliminate the "dead-end" nature of the non-academic tracks and facilitate upward mobility within the secondary education system. As a result, new school forms—highly diverse across federal states—have been introduced to serve as bridges toward higher educational qualifications. For example, a student may begin in lower secondary education, obtain an initial qualification, then transition to an intermediate school track to earn a second qualification, and finally move into the academic track or a related school type in upper secondary education (such as a Fachoberschule) to ultimately obtain the HEE. This layered structure means that, regardless of the track chosen after primary school, it is theoretically possible to attain the highest secondary school qualification—provided the student demonstrates sufficient academic performance and motivation to continue.

[1] For a detailed overview of how pupils are distributed across school tracks and which trajectories they take through secondary school refer to Henninges et al. (2019).

### 2.2. Diversion effects in secondary education

Students who are highly motivated at the beginning of secondary education and hold strong educational aspirations generally face two main options. They can either enter the academic track and obtain the desired final qualification (Abitur) after eight or nine years of schooling, depending on the federal state, or they can choose a lower school track and pursue sequential upgrading. The latter path is perceived as less risky, allowing students to hedge their investments—particularly in terms of time and effort—by first earning several intermediate, albeit lower, qualifications. In contrast, students who enter the academic track and later drop out—perhaps due to the more demanding academic standards—may find themselves without any formal qualification, thus prolonging the time required to obtain a first degree. This dynamic highlights why risk-averse students, even those with high aspirations, are more likely to choose non-academic pathways. Parents from socially disadvantaged backgrounds are often particularly hesitant to enroll their children in the academic track, as it generally takes longer to achieve a first qualification that enables entry into the labor market or the Dual System (Hartlaub & Schneider, 2012). However, due to the non-academically focused learning environments they enter with this decision, they may lose their aspirations at some point. There are multiple reasons why this might happen. Based on the literature, we identify three main sources of differences between academic and non-academic tracks (Bittmann & Schindler, 2021; Traini et al., 2021).

Firstly, since the curricula in nonacademic educational tracks are generally more oriented toward practical skills and vocational content rather than theoretical or academic subjects, it can be inferred that students are not actively encouraged to develop an interest in more scientific or scholarly fields (cf. Bayer, 2020, pp. 69–71). This practical focus may limit students' exposure to academic disciplines, thereby diminishing the likelihood that such interests are cultivated.

Secondly, the social background of students enrolled in nonacademic tracks tends to be characterized by families in which parents have not attained higher education and are typically employed in manual labor or lower- to mid-level service roles. This particular social composition increases the probability that students' educational and career ambitions are shaped by the values and expectations of their peer groups and close social contacts—often referred to as significant others—who may themselves lean toward nonacademic life paths (Kristen, 2002; Sewell et al., 2003).

Thirdly, student aspirations may also be shaped by the influence of teachers. As suggested by Van den Broeck et al. (2020), educators working in nonacademic school environments might find it challenging to foster academic ambition in students who show intellectual potential or interest, especially when the institutional demands of their teaching roles emphasize the delivery of practical competencies over academic enrichment.

Taken together, these three factors are expected to contribute to what is commonly referred to as school-track effects. In particular, the structure and social environment of non-academic tracks are likely to lead students—who may have initially aspired to obtain qualifications granting access to higher education—to revise their educational goals downward. At this point, it is important to emphasize that choosing a vocational track and learning a trade is not inherently a negative outcome. Skilled craftspeople are essential for the functioning of any society, and no society can consist entirely of academics. As long as the decision to pursue vocational education does not strongly depend on social background or initial track placement, overall levels of educational inequality remain unaffected.

### 2.3. Previous findings

Since the development of academic achievement is crucial for final educational outcomes, it is not surprising that many studies have examined whether achievement progresses differently across school

tracks—a phenomenon often referred to as scissor effects. However, a recent comprehensive review of more than 20 studies reports mixed findings, and no clear consensus has yet emerged regarding the existence or consistency of these scissor effects (Traini et al., 2021). To mention some of these highly influential studies, a widely cited contribution by Baumert et al. (2006) did not only characterize differentiated learning environments but also showed how relevant outcomes, such as cognitive ability, are not only stratified by initial track but also develop differently. These results have been confirmed using different datasets and outcome measures, such as reading ability (Becker et al., 2012; Retelsdorf et al., 2012). A quite recent study furthermore showed that the German system of tracking does not make initial social differences more pronounced but leads to increased academic performance, as originally intended by tracking (Esser & Seuring, 2020). However, these findings, in turn, have been criticized as well, resulting in a debate that is still ongoing (Lorenz et al., 2023; Esser, 2024).

Besides providing a broad review of the literature, the authors (Traini et al., 2021) also conduct their own analyses and find scissor effects when comparing the academic track to the intermediate track (*Realschule*) concerning math and reading achievement from grades 5–9. Note that this study does not focus on diversion effects, and the students are not strictly positively selected based on aspirations, even if a matching approach for other control variables is implemented. In general, diverging academic achievement can, at most, indicate differing educational outcomes, and these findings only serve as an indicator. Concrete research on diversion in secondary education in Germany needs to be more extensive.

Some previous studies report potential diversion effects; however, their precise impact remains unclear (Becker & Hecken, 2008; Müller & Pollak, 2004; Schindler, 2014). One newer study utilizes retrospective survey data and a simulation approach to gauge whether introducing new school forms since the 1960s has influenced educational inequalities (Schindler & Bittmann, 2021). The authors conclude that it is unlikely that these reforms have led to a decreasing social inequality concerning the attainment of higher educational eligibility or tertiary attainment. While these findings are relevant, they compare two rather old cohorts and cannot shed light on current developments and diversion in the German system. One newer study using data on more recent educational reforms concludes that de-tracking, reducing the number of school tracks in German upper secondary education, potentially increases social inequality, even if average aspirations rise (Holtmann et al., 2024). One highly relevant study with prospective student data investigates the influence of tracking on the development of educational aspirations between grades 5 and 9 (Bittmann & Schindler, 2021). The conclusions are that (both realistic and idealistic) educational aspirations drop drastically for students in the non-academic tracks but remain virtually constant in the academic track, even after implementing a rather strict sample selection and matching approach; the adjusted track difference amounts to about 17.4 percentage points in grade 9 (ibid: 246). These results underscore that relevant antecedents of educational outcomes diverge largely between tracks, which can potentially influence final educational outcomes. Finally, a German study tests whether cultural heterogeneity, which directly relates to the effects of tracking, influences the realization of initial aspirations negatively (Gerth, 2022). However, no strong evidence for this hypothesis is presented empirically. Another recent German study, investigating why some students enter the dual system after obtaining the higher education eligibility, concludes that risk-averse students might want to avoid a tight labour market and continue education instead (Hartung & Weßling, 2024). This result implies that diversion can also *decrease*, depending on the current labour market situation (especially in blue-collar occupations). Finding internationally comparable studies is difficult due to the rather distinct features of the German system. An exception is two Danish studies, indicating that disadvantaged students might react especially sensitively to adverse learning environments (Karlson, 2015, 2019).

While most studies so far have only considered diversion effects with respect to educational outcomes (test scores, grades, or degrees), the dimension of satisfaction and happiness should not be overlooked. As important as these educational aspects are, there is more to life than a good performance. If certain tracks or pathways systematically lower the happiness or life satisfaction of students, this might be concerning. It is known that life satisfaction tends to decrease over the course of secondary education (Hascher & Hagenauer, 2011) and trajectories by track are not identical. At the start of grade five, students in the academic track have higher satisfaction values than students in the non-academic tracks, however, both tracks then shown a rather steep decline over the following years to converge on the same level (Herke et al., 2019; Wu & Becker, 2023). While these findings concern the general student population, the question arises whether the sample of students that experience diversion effects show the same trajectories.

In summary, the available studies outline plenty of indicators for the effects of diversion in German secondary education. However, since most studies do not utilize relevant outcomes, such as the final educational qualification of secondary education, it remains to be seen whether tracking's strong and long-lasting negative effects materialize. We attempt to fill this research gap by utilizing recent high-quality longitudinal data. By having complete educational trajectories in secondary education available, including many relevant control variables, we can quantify the actual magnitude of the diversion effects.

## 3. Materials and methods

### 3.1. Data and sample

We utilize German National Educational Panel Study (NEPS) data to answer our research questions.[2] The NEPS is the most ambitious German research project to study the role of education throughout life. We employ starting cohort 3 (SC3), which sampled students directly after they transitioned to secondary education in grade 5. Due to the long-running panel of more than ten years, the students are older than 20 years in the most recent survey wave and have left secondary education. This means we can observe complete educational trajectories for these students through secondary education, and no right-censoring should occur. While it is possible in the German system to obtain further educational qualifications as an adult (second chance education; *Zweiter Bildungsweg*), this happens rarely, and we can be confident that the period of SC3 is sufficient to observe complete educational trajectories for most students. Regardless of which pathway these students have taken (entered the labour market, the vocational system, or tertiary education), their highest secondary education qualification has been surveyed. This enables us to observe relevant outcomes.

In SC3, secondary schools all over Germany were randomly sampled, and within each sampled school, two grade-5 classes were randomly selected. All students and their parents were invited to participate in the NEPS. This design benefits our research agenda since it includes a wide range of students from many regions in Germany. Since the initial survey in grade 5 was conducted rather soon after the transition occurred (between 2 and 7 months after transitioning), it is unlikely that the system of tracking in secondary education had large effects on the students in wave 1 of the survey, which means that the occurrence of reversed causality is unlikely.

For our analyses, we restrict the sample to fit our needs. 5778 students have participated in wave 1 of NEPS SC3. We focus on this group of students who have participated from the beginning and exclude any refreshment samples since relevant information for their early secondary education is unavailable. We exclude all students who have not

[2] This paper uses data from the National Educational Panel Study (Blossfeld & Roßbach, 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network.

transferred to a regular secondary school in grade 5, which removes all students in special needs schools (*Förderschulen*). Note that this selection also removes students from Berlin and Brandenburg since, in these federal states, the selection for secondary education does not occur after grade four but later. This leaves 4898 students in the sample. Finally, we want to focus on a sample of highly motivated families, which we define as having uniformly high educational aspirations. We only retain students where both the students and also their parents have reported that they would like to obtain higher education eligibility (HEE, *Abitur*). Note that parents and students usually agree on their aspirations as this is the case for 81 % of the families in the dataset. We refer to idealistic educational aspirations for this purpose.[3] By doing so, we select a sample where the families agree that the *Abitur* would be the desired educational outcome regardless of any constraints. This reduces the sample size to 1966. While this is a drastic reduction, it is necessary, since only students with initially high aspirations can be affected by diversion, that is, a loss of aspirations. Students who never report that they want to obtain a high educational qualification are, hence, not part of this study. By applying a rather strict definition (parents and students need to express high aspirations), we believe that our effects will present a lower limit to effect sizes as, for example, students with initial high aspirations, who lack parental support, might lose the aspirations much more easily. Consequently, if highly positively selected students lose their aspirations due to tracking, students with less firm aspirations might be even more likely to do so. The number of students available for the following analyses will be slightly reduced to about 1938 cases due to excluding students not in the region of common support; this aspect refers to the statistical modelling and is explained in more detail in the next section.

### 3.2. Strategy of analysis

One of the main goals of our study is to assess the extent to which different school tracks in secondary education affect highly motivated students. To do this, it is essential to distinguish between selection effects and the causal effects exerted by the tracks themselves. While the German tracking system is explicitly designed to sort students based on academic ability, it also allows for sequential educational upgrading, as outlined in the theoretical background above. By minimizing selection effects as much as possible, we aim to estimate the residual impact attributable to differences between tracks. To achieve this, we account for the factors that influence track selection. In the German system, the primary criterion for selection is academic achievement. Based on students' grades in the final year of primary school, teachers issue a recommendation for secondary school placement. Depending on the federal state, this recommendation may be binding or may be overridden by parents (although choosing a less demanding track than recommended is always permitted in all states). We seek to measure both objective academic achievement and subjective influences, which can significantly shape parental decision-making. A second major factor influencing track choice is the family's social background. It is well established that socially advantaged and highly educated parents are more likely to choose the academic track for their children, even when academic performance does not fully support this decision (Bittmann, 2021). These parents value education highly and are convinced of the benefits of the academic track, even if other pathways to the *Abitur* are possible. Consequently, we measure multiple aspects of social origin (financial means, parental education levels, and social status through occupations in the labour market). Finally, there are a few other aspects we control for, such as gender or age of the student and also some structural influences. By measuring all these various factors at the very start of secondary education, we are confident that they explain selection into tracks and partial out these effects.

As there are exactly two groups at the start of secondary education in our design (those studens who have transferred to the academic track and those who have not), we would like to make these two groups as equal as possible concerning all control variables to have a fair comparison of what tracking does to students afterwards. We utilize entropy balancing (EB), a weighting approach, to achieve this (Hainmueller, 2012). EB considers all control variables and generates weights for each observation so that the two groups are highly similar regarding the controls. The advantage of EB is, in contrast to other popular methods, such as regression analyses, that the balancing of all variables can be inspected. By doing so, we can demonstrate statistically that the two groups are comparable concerning all control variables. If observations cannot be balanced as no counterpart in the other treatment group can be found, they fall out of the region of common support and are removed from the analyses. This is also beneficial since this step removes observations that could bias the results yet would go unnoticed in regression models. A third advantage of this matching approach is that the functional form between outcome and treatment is irrelevant. Most regression models assume a linear relationship, which must be tested and potentially accounted for if violated. In contrast, the matching does not impose such a relationship. Further below, we empirically demonstrate that the two groups of interest are indeed highly similar after applying the EB weights, which balance means and standard deviations. Balanced are time-constant control variables and time-varying outcome variables at the first measurement point (grade 5). By doing so, we can account for any pre-tracking differences and study in detail whether trajectories are different throughout secondary education. It is crucial to understand that this balancing forces the group means of all time-varying outcome variables to be equal at the first measurement point (grade 5) yet allows different trajectories afterwards.

We compute panel regression models for the time-varying outcomes (math and reading achievement, life and school satisfaction) with a single independent variable (the group indicator) and the EB weight. These models incorporate random intercepts for each student, which is necessary to account for the dependencies in the data (observations nested within students). All these outcomes are continuous; hence, we estimate linear (OLS) models. For the outcome of the educational degree, which is binary, we compute a linear probability model (LPM). Technically, this is also in an OLS regression model. Since the only independent variable is binary, this is fine and lets us interpret the regression coefficient as a percentage change (Angrist & Pischke, 2009; Wooldridge, 2010).

We utilize multiple imputations with chained equations (MICE) to account for missing data (Allison, 2001). This widespread and statistically well-research approach enables us to combat selective dropout and attrition, which is common in long-running studies. By using all available information in the data, the approach estimates plausible values for missing data. By doing so, we avoid bias due to selective dropout. If we assume that the most unhappy students (potentially also with the worst outcomes) refuse to participate after some waves, a listwise estimation approach could result in wrong conclusions. MICE uses cross-correlations to estimate missing values and prevents such bias as much as possible. We utilize predictive mean matching for all continuous and binary variables and ordered logistic models for the parental education level, number of books at home, and town size. We generate 40 imputed datasets and check the convergence of the results graphically and numerically. By doing so, we can test that only sensible values are generated and that the distribution of generated values is similar to observed values. All computations are done in Stata 16.1. Entropy

[3] The parental item is: "Even though it will take quite a while, this is in regard to your wishes and expectations for the school-leaving qualifications of <name of target child> . No matter how good <name of target child> 's achievements are: what school-leaving qualification would you like for him?" The filial item is: "Regardless of which school you attend and how good your grades are, what kind of school-leaving qualification would you like to obtain?"

balancing weights are computed using *kmatch* (Jann, 2017). To create some figures, we use the additional package *mimrgns* (Klein, 2014). Regression sensitivity checks are computed with *regsensitivity*.[4]

### 3.3. Operationalization

Tracking is operationalized as a binary variable. If a student has transferred to the academic school track (*Gymnasium*) after primary schooling, this is coded as starting in the academic track. This also includes students in the academic track in cooperative comprehensive schools. If any other school track is selected (*Hauptschule*, *Realschule*, *Gesamtschule*), this is counted as a non-academic schooling form. We employ this binary coding scheme for theoretical reasons (as explained above since all non-academic tracks require another educational decision). Our data need to be larger to utilize a finer coding scheme, as some groups would become extremely small since only a few dozen cases are available in tracks such as the *Hauptschule*.

The first dependent variable of the analyses, the highest educational qualification, holds significant importance in our research. It reflects the culmination of a student's secondary education journey. In the later waves of SC3, students are repeatedly surveyed about which types of schools they are currently in and what degrees they want to obtain or have obtained. By using these data, we can check the highest degree that has ever been obtained. If this is the HEE (*Abitur*), we code this as 1. If any lower degree or no degree at all has been achieved, we code this as 0. This binary coding scheme directly refers to the aspiration variable used in the selection, as explained above. We can consider students who have obtained the *Abitur* as being in line with their initial educational aspirations; all others have either abandoned these high aspirations or failed to achieve them. Due to the nature of this variable, it is only available at the very end of secondary education and cannot be traced as the students progress through school. To gain more insight, we have selected four other variables that are relevant for academic success but also overall happiness.

To measure the development of academic achievement over time, we use the comprehensive NEPS achievement tests conducted four times (in grades 5, 7, 9, and 12). These tests are conducted for mathematics and reading achievement and are relevant to judge how the students develop their skills over time. In contrast to grades, these tests are highly standardized and well-suited for both within- and between-student comparisons. The tests are identical for all tracks and are not influenced by the teacher assessment (removing any tertiary effects). In contrast to grades, which only contain six levels in Germany, the test scores are normally distributed and offer many values for a nuanced judgment. These reading and math competencies were assessed using paper-based achievement tests designed by the NEPS. These tests were constructed based on established frameworks in large-scale assessments like the Programme for International Student Assessment (PISA). A consistent scaling procedure was applied to generate unidimensional proficiency scores (Pohl & Carstensen, 2013). To facilitate meaningful comparisons across grade levels, the tests were linked to a common scale (Fischer et al., 2016). The reading assessments were conducted across different grade levels. Grade 5 tests included 32 items, Grade 7 had 40 items, Grade 9 had 46 items, and Grade 12 had 29 items. To enhance precision, the Grades 7, 9, and 12 tests followed a branched design, adjusting difficulty based on students' prior performance. These assessments covered different text types (information, instruction, advertising, commenting, and literary texts) and addressed three cognitive requirements. No common items were shared across grades to avoid memory effects, and an anchor-group design linked the tests across measurement occasions. Proficiency scores demonstrated satisfactory reliability in each grade (0.77, 0.83, 0.81, and 0.80, from grades 5–12) (Gnambs & Lockl, 2023).

Mathematical competency was assessed similarly. Grade 5 had 25 items, Grade 7 had 23 items, Grade 9 had 34 items, and Grade 12 had 30 items. Grades 9 and 12 tests followed a branched design, offering booklets with varying difficulty levels. These assessments covered content areas (quantity, space and shape, change and relationship, and data and chance) and required six cognitive components for successful task solutions. Some common items were shared across successive grades, and an anchor-item design linked the mathematics tests. Proficiency scores demonstrated good reliability at all measurement points (0.80, 0.76, 0.81, and 0.77) (Gnambs & Lockl, 2023). Note that both competence tests were carefully developed and pre-tested extensively. The NEPS institute provides ample information on the design, scaling, and statistical quality and validity of these tests in various publications.[5]

Finally, we would also like to see how well students feel about their progress through secondary education. To do so, we utilize two variables that measure subjective satisfaction. The first is overall life satisfaction: "How satisfied are you currently with your life?". This item is taken from the *TNS infratest Sozialforschung* survey as an established item.[6] The second item measures satisfaction with current schooling: "How satisfied are you with your situation at school?". This item is an adaption developed by the NEPS to measure satisfaction of students in the school setting. Both items are measured on an 11-point scale with values ranging from 0 ("completely unsatisfied") to 10 ("completely satisfied"). While academic success is relevant to one's life, checking whether students are happy and satisfied with their lives and the current situation is also relevant. These two items enable us to track these statistics over time as they are surveyed in every wave of the NEPS.

To account for confounding, we utilize a wide range of control variables. For objective academic achievement and overall cognitive ability, the NEPS provides four variables measured in grade 5: math achievement score, reading achievement score, and two measurements of overall cognitive ability (DGCF, Domain-General Cognitive Functions). For subjective academic performance, we have two measurements: first, the parents reported in their survey whether the student ever had to repeat a school year in primary schooling, which is a binary variable. Second, the parents also report whether the teacher in the final year of primary schooling has given the child a recommendation for the academic track. These variables indicate how well the student did in primary schooling and are highly relevant for the track selection in secondary education.

Next, we measure the social origin using multiple variables. First, we have the (logged) total post-tax household income, adjusted by OECD definition for the number of individuals in the household. Second, we use the information on the highest parental education for both mother and father with four levels: no degree or lower degree (*Hauptschulabschluss*), intermediate degree (*Mittlere Reife*), higher education eligibility (*Abitur*), any tertiary degree. We also have the current ISEI of both parents, which is based on the occupation parents have in the labour market. The ISEI ranges from 11 (cleaner) to 80 (judge), and higher values indicate a higher social status. In addition to these central measures, we also have further variables related to the family of the student, such as whether the student lives in a nuclear family with both parents or with a single parent. Migration background is measured with four levels: if the student and both parents are born in Germany, this is counted as native. Other constellations are: student born abroad, one parent born abroad, both parents born abroad. We also include the number of siblings the student has as well as the number of books at home (measured using a variable with six categories ("0–10 books" up to "more than 500 books"). We control for student gender and age

[4] https://github.com/mattmasten/regsensitivity

[5] A package containing relevant information for all competence tests and survey waves is available from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC3/9-0-0/SC3_9-0-0_C.zip (2024–06–20).

[6] https://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.356304.de/soepfrabo_jugend_2009.pdf (2024–06–13)

(computed for January 1, 2011).

Finally, we account for structural differences between schools and educational systems. The first is the size of the town in which the school in grade five is located. This variable has three levels: up to 20,000 inhabitants, 20,000 to 100,000 inhabitants, and more than 100,000 inhabitants. The second is whether the type of track recommendation given by the class teacher in elementary school is binding or not. At the time of the survey when this was relevant (school year 2010/11), the following federal states still had this bindingness active: Bayern, Baden-Württemberg, Brandenburg, Saxony, and Thuringia. In these states, the decision for a school track was hence partially pre-determined by what the teacher recommended, and overruling this decision was rather difficult. We argue that this is the single most relevant structural factor for our analysis as it greatly limits the free choice of parents. Hence, we not only include the type of recommendation as a control variable but also conduct further robustness checks for a sample that only includes federal states where this binding character has already been abolished (see below). Summarized, we argue that the control variables selected explain (self-) selection into tracks well, which is based on our research results (Bittmann, 2021, 2023; Hillmert & Jacob, 2010). By accounting for these, a fair comparison of trajectories in different school tracks is facilitated. Still, we would like to point out that our results cannot compute pure causal effects. While we are confident that our control variables can explain self-selection into tracks to a large degree, other influences can never be ruled out as we are using observational data. Researchers should be aware of this fact when interpreting our findings. We also compute additional sensitivity checks below to assess the potential influence of unobserved confounders.

## 4. Results

### 4.1. Descriptive findings

We start by presenting a table containing basic descriptive information. Of the 1938 students in the analytical sample, 77.2 % have transitioned to the academic track in grade 5, and the remaining 22.8 % have not. This rather large imbalance already hints at a clear preference for the academic track in families with high aspirations. We also see further differences between the two groups in Table 1. Regarding the outcome of receiving the HEE, we see that 61 % of those in the non-academic track have obtained this degree, but 88 % have in the academic track, which is a raw difference of about 27 percentage points. Considering the background variables, in the group of students who have transitioned to the academic track, 91 % have received a recommendation for this track, 82 % are natives, and 13 % are raised in a single household. In contrast, for students in the non-academic tracks, only 43 % have received a recommendation, 75 % are natives, and 24 % live in a single household. This general advantage for the students in the

**Table 1**
Descriptive comparison of both school tracks before and after matching.

| | Pre-matching (raw) | | | | Post-matching | | | |
|---|---|---|---|---|---|---|---|---|
| | Other tracks (T1) | | Academic track (T2) | | Other tracks (T3) | | Academic track (T4) | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Ever obtained HEE | 0.61 | 0.49 | 0.88 | 0.33 | 0.70 | 0.46 | 0.84 | 0.37 |
| Math achievement score* | 0.054 | 0.97 | 0.72 | 0.95 | 0.55 | 1.01 | 0.55 | 1.01 |
| Reading achievement score* | 0.045 | 1.19 | 0.67 | 1.09 | 0.53 | 1.14 | 0.53 | 1.14 |
| DGCF1 | 45.5 | 14.3 | 45.7 | 12.9 | 45.6 | 13.2 | 45.6 | 13.2 |
| DGCF2 | 6.93 | 2.48 | 8.03 | 2.19 | 7.79 | 2.30 | 7.79 | 2.30 |
| Repeated a grade | 0.063 | 0.24 | 0.012 | 0.11 | 0.023 | 0.15 | 0.023 | 0.15 |
| Teacher recommendation | 0.43 | 0.49 | 0.91 | 0.29 | 0.80 | 0.40 | 0.80 | 0.40 |
| Log. equivalent HH income | 7.31 | 0.43 | 7.46 | 0.41 | 7.42 | 0.42 | 7.42 | 0.42 |
| ISEI of father | 49.8 | 18.7 | 54.6 | 18.0 | 53.6 | 18.2 | 53.6 | 18.2 |
| ISEI of mother | 48.1 | 15.0 | 52.6 | 15.3 | 51.7 | 15.3 | 51.7 | 15.3 |
| Education mother | | | | | | | | |
| Low | 0.14 | 0.35 | 0.064 | 0.24 | 0.081 | 0.27 | 0.081 | 0.27 |
| Intermediate | 0.41 | 0.49 | 0.35 | 0.48 | 0.36 | 0.48 | 0.36 | 0.48 |
| HEE | 0.22 | 0.42 | 0.29 | 0.45 | 0.28 | 0.45 | 0.28 | 0.45 |
| Tertiary | 0.22 | 0.41 | 0.29 | 0.46 | 0.28 | 0.45 | 0.28 | 0.45 |
| Education father | | | | | | | | |
| Low | 0.25 | 0.43 | 0.14 | 0.35 | 0.16 | 0.37 | 0.16 | 0.37 |
| Intermediate | 0.27 | 0.44 | 0.25 | 0.43 | 0.25 | 0.43 | 0.25 | 0.43 |
| HEE | 0.18 | 0.39 | 0.19 | 0.39 | 0.19 | 0.39 | 0.19 | 0.39 |
| Tertiary | 0.29 | 0.46 | 0.42 | 0.49 | 0.40 | 0.49 | 0.40 | 0.49 |
| Single parent | 0.24 | 0.43 | 0.13 | 0.34 | 0.15 | 0.36 | 0.15 | 0.36 |
| Migrant status | | | | | | | | |
| Native | 0.75 | 0.43 | 0.82 | 0.39 | 0.80 | 0.40 | 0.80 | 0.40 |
| Born abroad | 0.041 | 0.20 | 0.025 | 0.16 | 0.028 | 0.17 | 0.028 | 0.17 |
| One parent born abroad | 0.12 | 0.32 | 0.10 | 0.30 | 0.10 | 0.31 | 0.10 | 0.31 |
| Both parents born abroad | 0.090 | 0.29 | 0.056 | 0.23 | 0.063 | 0.24 | 0.063 | 0.24 |
| Size of town of school | | | | | | | | |
| Up to 20k | 0.32 | 0.47 | 0.26 | 0.44 | 0.28 | 0.45 | 0.28 | 0.45 |
| 20–100k | 0.43 | 0.50 | 0.45 | 0.50 | 0.45 | 0.50 | 0.45 | 0.50 |
| More than 100k | 0.25 | 0.43 | 0.28 | 0.45 | 0.28 | 0.45 | 0.28 | 0.45 |
| Number of books at home | 4.16 | 1.23 | 4.58 | 1.14 | 4.48 | 1.17 | 4.48 | 1.17 |
| Number of siblings | 1.63 | 1.15 | 1.46 | 1.02 | 1.50 | 1.05 | 1.50 | 1.05 |
| Age in January 2011 | 11.0 | 0.40 | 10.9 | 0.36 | 10.9 | 0.38 | 10.9 | 0.38 |
| Female student | 0.54 | 0.50 | 0.49 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Binding teacher recommend. | 0.24 | 0.43 | 0.42 | 0.49 | 0.38 | 0.49 | 0.38 | 0.49 |
| Life satisfaction* | 8.17 | 2.28 | 8.55 | 1.93 | 8.47 | 2.02 | 8.47 | 2.02 |
| School satisfaction* | 7.79 | 2.31 | 8.23 | 2.07 | 8.13 | 2.13 | 8.13 | 2.13 |
| Observations | 429 | | 1537 | | 429 | | 1537 | |

Source: NEPS SC3, imputed data (M = 40). Variables marked with an asterisk are time-varying, reported is the value from grade 5 (wave 1 of the survey). HEE = higher education eligibility, HH = household, ISEI = International Socio-Economic Index of Occupational Status, DGCF = Domain-General Cognitive Functions.

academic track continues for variables that measure academic performance or social origin. This fact is well-known and explains how selection in the German system works. Without considering these pre-tracking differences, one cannot explain how the different school tracks influence the students further, as these initial differences are huge and will not vanish. Fortunately, we can approach this issue with statistical balancing. After generating and applying the balancing weights to the summary statistics, the groups are highly similar (columns T3 and T4). All control variables are virtually identical, with the standard deviations as well. In contrast to regression approaches, which do not allow the creation and inspection of these values, we can demonstrate numerically that the balancing approach does work and highly comparable groups are generated. Note that students outside the range of common support have already been removed, so the number of observations is constant throughout the entire table (and all of the following analyses). The propensity scores of both groups are presented in the appendix (Figure A1).

### 4.2. Longitudinal developments

In this section, we demonstrate how achievement in math and reading and student satisfaction develop from grade five to grade 12, covering seven school years. As described above, we first generate an entropy balancing weight for grade five with all control and dependent variables. This guarantees that the four outcome variables of interest are identical at the start of secondary education and that any divergent developments between the academic and the non-academic tracks are a substantive tracking effect and not due to selection alone. After generating the weight, we estimate the panel-regression models with the track as the sole independent variable and include the weights. For a convenient interpretation, we first generate graphs to trace the development over time. Achievement scores are based on the NEPS tests and available for survey school grades 5, 7, 9, and 12. Satisfaction scores are available for all school grades from 5 to 12. The results are shown in Fig. 1. In contrast to classical (parametric) growth-curve models, our approach is highly flexible. By introducing interaction effects between the time variable and the outcome, we do not impose a pre-defined functional form (e.g. linear, sigmoid, …) but allow a more data-driven estimation to account for potential non-linearities. This facilitates a graphical inspection of the results.

As we can see, the outcomes are virtually identical for both groups in grade 5, which means that the weighting approach works as intended. Any divergent development in the subsequent survey wave is not due to any initial differences but reflects tracking effects. Regarding the math score (upper left graph), the average achievement increases over time, which is how the NEPS tests are designed to account for increasing ability. However, there are virtually no group differences as the two lines are very close, and the 95 % confidence intervals always largely overlap. This means the trajectories are parallel, accounting for initial differences at the start of secondary schooling. In other words, the development of math achievement is similar, regardless of whether a student has transferred to the academic or any other school track. For reading achievement (upper right graph), the same conclusions hold as the two lines are always close, and no divergent development is visible. We also note that no divergent developments are present for the two satisfaction measures. Students report similar satisfaction values, regardless of whether they attend the academic track or any other school track. As an additional numerical check, we have provided tables for the models (Table 2). In these, we utilize time as a continuous variable, which seems reasonable given the visual inspection in Fig. 1 does not show strong non-linear trends. The track-time interaction term is relevant to tell whether the average development over all available survey waves differs between tracks. As the coefficients outline, no such interaction term is statistically significant on the 5 % level. To summarize, the trajectories for all four outcome variables are similar for both tracks. Achievement and satisfaction development were similar for all students in the sample, regardless of which track they have transferred to, after accounting for initial differences and selection effects using the entropy balancing approach. Note that the intra-class correlation indicates that the values within each student are highly correlated, which is logical since especially achievement is dependent on some time-constant traits or individual talent that does not change strongly over time.

### 4.3. Probability to obtain higher education eligibility

Finally, we would like to investigate how the probability of obtaining higher education eligibility differs by track placement in grade five. We compute three different models for more insight and robustness (see Table 3). The first model is the raw difference between tracks without any controls or matching. This model gives the overall effect of tracking when selection is not accounted for (and is basically identical to the difference reported in Table 1). The difference amounts to 26.8 percentage points, which means that students who have transferred to the academic track after primary schooling are almost 27 percentage points more likely to obtain the *Abitur* than those who have transferred to any other school track. This denotes a rather large advantage for the academic track, however, is an expected result given that selection effects are unaccounted for. Hence, this is not the "causal" effect of tracking alone. Next, we try to account for the initial differences between the tracks using the balancing approach. However, please keep in mind that this approach usually cannot recover pure causal effects as no survey can contain all potential confounders. By adding this weight, we adjust for all control variables and initial differences in the other outcome variables (achievement and satisfaction). As soon as these effects are considered, the difference between the two tracks is reduced to 13.3 percentage points. This means the difference between the two tracks is more halved when selection effects are partialed out statistically. Finally, as a robustness check, we estimate an OLS regression model where we do not utilize the generated balancing weight but add the control variables directly (using robust $HC_3$ standard errors). This model shows a difference of 11.5 percentage points between tracks. As these numbers are similar to the balancing approach, we can conclude that the chosen method does not influence the findings significantly. Both approaches are valid and give sensible results. We would argue that the matching logic is more suitable for our study design and regard it as the main result. Additionally, track placement alone explains only about 8.2 % of the outcome variable. This share doubles to more than 19 % when adding all control variables. Still, more than 80 % of the variation in the outcome is unaccounted for, even when using many relevant explanatory variables. Since the time frame is rather large (approximately ten years), it is obvious that the outcome is affected by many more influences that can potentially work over the entire period. While our model is fine regarding our explanatory and theoretical framework, purely predicting outcomes requires much more data from recent survey waves.

### 4.4. Robustness checks

As already outlined before, we would like to strengthen the robustness of our results by conducting further checks. We have implemented three relevant ones: First, testing whether the results hold when restricting the sample to federal states where no binding teacher recommendation is present; second, testing whether the results also hold when *realistic* instead of idealistic aspirations are investigated; and third, testing how results change when the presence of unobserved confounders is simulated, following the approaches of Diegert et al. (2022) and Oster (2019). Fourth and final, we also test the influence of extreme entropy balancing weights on the diversion results (weight trimming). Summarized, the results are quite robust and tests one to three hardly change the results at all. Regarding the extreme entropy balancing weights, we notice that the diversion effect tends to grow when rather

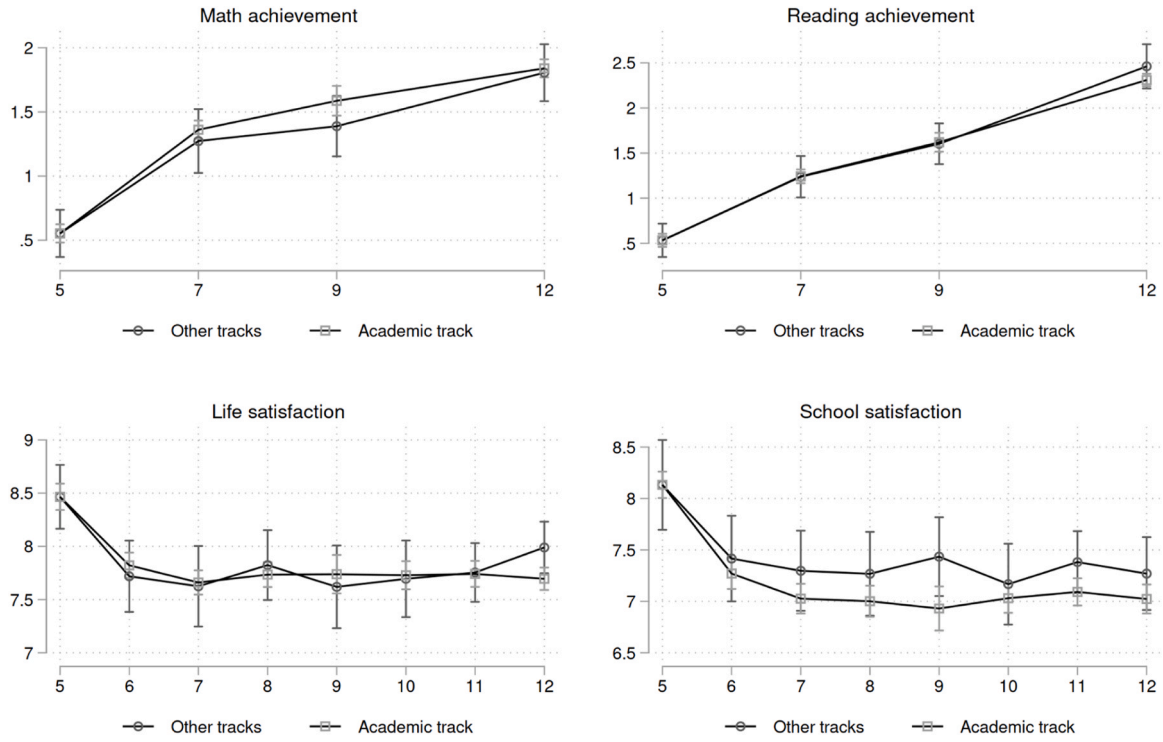


**Fig. 1.** Trajectories of outcome variables by track. Source: NEPS SC3, imputed data (N = 1938, M = 40). 95 % confidence intervals included, entropy balancing weights applied, only region of common support included.

**Table 2**
Numerical results for the growth-curve models.

| | Math achievement | Reading achievement | Life satisfaction | School satisfaction |
|---|---|---|---|---|
| Academic track | 0.0290 | 0.141 | 0.234 | 0.00121 |
| | [ 0.271,0.329] | [ 0.214,0.496] | [ 0.214,0.681] | [ 0.671,0.669] |
| Wave | 0.165*** | 0.267*** | 0.0374 | 0.0766** |
| | [0.140,0.191] | [0.231,0.304] | [ 0.0799,0.00501] | [ 0.133, 0.0203] |
| Academic track X Wave | 0.00617 | 0.0210 | 0.0290 | 0.0272 |
| | [ 0.0202,0.0325] | [ 0.0599,0.0179] | [ 0.0765,0.0185] | [ 0.0881,0.0337] |
| Constant | 0.109 | 0.747*** | 8.155*** | 8.072*** |
| | [ 0.395,0.177] | [ 1.086, 0.409] | [7.744,8.566] | [7.441,8.702] |
| Individuals | 1938 | 1938 | 1938 | 1938 |
| Observations | 7752 | 7752 | 15,504 | 15,504 |
| ICC | 0.543 | 0.476 | 0.402 | 0.400 |

Source: NEPS SC3, imputed data (M = 40). 95 % confidence intervals in brackets, entropy balancing weights applied, only region of common support included. * p < 0.05, ** p < 0.01, *** p < 0.001

**Table 3**
Outcomes regarding achieving higher education eligibility.

| | **Raw (Unadjusted)** | **Matching** | **Regression** |
|---|---|---|---|
| Transition to the academic track | 0.268*** [0.206; 0.329] | 0.133** [0.038; 0.228] | 0.115** [0.044; 0.186] |
| Adj. $R^2$ | 0.082 | - | 0.191 |
| Observations | 1938 | 1938 | 1938 |

Source: NEPS SC3, imputed data (M = 40). 95 % CIs in brackets, only region of common support included. * p < 0.05, ** p < 0.01, *** p < 0.001

extreme weights are removed. This could mean that a rather special group of students has an over-proportional influence on the findings and diversion might be stronger than reported in the main model. However, when assessing these extra results, keep in mind that the main model only regards cases in the region of common support, meaning that comparisons of very extreme groups is already ruled out. For details see the appendix where all further analyses are presented and discussed due to space constraints.

## 5. Discussion and limitations

In contrast to earlier studies that primarily focus on intermediate outcomes during secondary education, our findings are somewhat more optimistic. With respect to the development of competencies and satisfaction, we find no significant differences in trajectories between the two track types after adjusting for self-selection—an outcome that slightly diverges from previous research (Traini et al., 2021).[7] While our results do replicate the well-documented overall decline in satisfaction scores over the course of secondary education, the trajectories for students in both academic and non-academic tracks are virtually identical. This is an encouraging finding, suggesting that highly motivated

[7] We assume that the difference lies in the different sample selection as the other paper does not select on initial aspirations.

students can successfully progress through the system—even when beginning on a non-academic track—without a decline in either competencies or satisfaction. For these students, attaining the highest secondary school qualification (Abitur) remains a realistic goal, and there is no evidence of a widening competency gap. These results support the notion of permeability and the possibility of continuous educational upgrading. However, when examining the final outcome—whether the Abitur was obtained—we observe a difference of approximately 13 percentage points between tracks. This is a substantial and meaningful difference that must be acknowledged, as it indicates that a notable share of initially highly motivated students were ultimately unable to achieve their educational goals, even after adjusting for self-selection. Some of our robustness checks suggest that this gap may be even larger in specific scenarios. Given that our analyses rely on counterfactual "what-if" scenarios—implemented through both regression and balancing methods—we must make several statistical assumptions. Under these assumptions, the "true" causal effect likely lies somewhere between the unadjusted difference of 27 percentage points and the adjusted estimate of about 13 percentage points produced by the entropy balancing (EB) model. By conducting a range of robustness checks, we enable readers to better assess the range of plausible outcomes. Even if the true difference is as low as 13 percentage points, this still represents a substantial loss of talent and potential. On a more positive note, these diversion effects appear to be smaller than previously estimated. For example, one study using a comparable sample reported a diversion effect of around 17 percentage points by grade 9, even after adjusting for selection (Bittmann & Schindler, 2021, p. 246). Especially since our data include a longer time frame, one could have suspected an even larger diversion. Some students may regain their aspirations that got lost at some point in secondary education, after obtaining a first educational degree (usually after grade 9 or 10) when they are forced to decide between entering the labour market or continuing education. Another reason diversion might *decrease in* a longer time frame (looking at 12 instead of only 9 school years) is that some students might try to avoid a tight labour market and continue education instead of entering the dual system (Hartung & Weßling, 2024). In any case, the diversion effects we can detect should be investigated in more detail in the future. As longitudinal data are available, it would be highly relevant to determine when exactly diversion is created and how this is happening. Is this a process that develops continuously over time (as some previous publications indicate), or are there "breakpoints" when one lower educational degree is achieved, and students must actively decide for or against continuing their educational journeys? Answering these questions is probably the only way to develop countermeasures and give practical advice to policymakers. One possibility is that it is not a deficit in competencies but in information that prevents students from continuing their educational journeys. As the German system is rather complex, especially in upper secondary education, since many different school forms exist, it would be desirable to support high-performing students in the non-academic tracks to navigate this system to enable them to achieve their initial goals. In this respect, especially teachers should be informed and encouraged to support and guide them.

Finally, we would like to discuss the limitations of the study. First, the sample does not perfectly represent the overall German student population, as some federal states are not included due to the sample selection (Berlin and Brandenburg). Some other states only have a few dozen students included. A larger sample would be required to account for these issues. Furthermore, we cannot perfectly match different educational systems and compare them rigorously, as the sample size is too low. Currently, the analyses compute averages over all federal states and school systems but heterogeneous effects cannot be explored. We attempt to combat this issue with the robustness checks we have provided to some extent, where we selected the most relevant structural differences. However, as in Germany, educational systems are different by state and not centrally directed by the federal government, focussing on these structural and institutional differences might be relevant in future studies (with larger and more detailed surveys). Second, our study is agnostic about the specific individual trajectories students take. We select students based on the track they attend directly after primary schooling; however, further changes have yet to be investigated. A student could directly switch after a single school year, either up or downgrading the track. However, switching tracks is a rather rare occurrence in the data. 85.8 % of all students stay on the same track in the first four years of secondary education, which hints for a rather large temporal stability. The exact amount of track switching is difficult to answer since some students have missing values on their tracking variable after grade five, making it uncertain whether they have switched tracks. For the goals of your study, this knowledge is optional as we are interested in final outcomes. Also, over ten years, many hundreds of possible switching patterns can arise, making even classifying patterns difficult.[8] For the aims of this study, we are not interested in how exactly students behave and which tracks they choose. A manifold of various school forms exists, especially after completing a first lower degree, which differs strongly by the federal state. With the available data, tracing these trajectories in detail is impossible. Our main objective is to see how the initial decision after primary schooling influences the outcomes, regardless of subsequent decisions.

Third, our observation window is ten years after the start of secondary schooling, meaning that students are approximately 21 years old. Even when a student had to repeat a school grade or switch tracks repeatedly, most students have completed secondary education at this age. However, our window does not extend to any form of second-chance education, normally taken up when individuals have entered the labour market. While we argue that our observation window is long enough to observe almost all trajectories, there might always be singular cases of students who take longer and are, hence, separate from the study due to right-censoring. Fourth, while it would be highly interesting to study sub-groups and especially how social origin and attendance of a non-academic school track interact, this is difficult due to the low number of students available. With less than 500 students in non-academic tracks, looking at specific sub-groups renders statistical power very low. Our tests have shown that such analyses, which also need to consider the other control variables, result in volatile results. Fifth, we emphasize that our results are only valid for highly motivated families, a strongly positively selected sample. However, this limitation refers to our main research question: whether diversion occurs. By definition, this only concerns students who initially hold high aspirations. Many students, especially those entering non-academic tracks, were never interested in obtaining the *Abitur*, which is also fine. One should never forget that the German system offers a wide range of relevant qualifications, leading to many potential occupations in the labour market, which can benefit and reward both the individual and society. We want to stress that the goal should never be to lead all students to the *Abitur* or even higher education. It should be possible for motivated students to fulfil their goals, regardless of which school track they start their educational journey on.

Sixth and final, the longitudinal analyses test track differences in reading and math, which is only a selection of all subjects taught at school. While these two are core subjects, which were also investigated in previous publications (Traini et al., 2021) and are taught at every track and grade, results might be different for other subjects. In this regard, the present analyses only show a part of the complete picture when specific skills or subjects are of interest and testing more subjects in future studies might be desirable.

## 6. Conclusion

Our results indicate that only weak diversion processes are present in

[8] For further insight on this issue regarding our data refer to a NEPS report (https://www.lifbi.de/Portals/2/Working%20Papers/WP_LXXXIII.pdf).

contemporary German secondary education—at least when examining a highly positively selected sample, namely families with generally high educational aspirations. After adjusting for a wide range of factors related to (self-)selection, the difference in final educational success between the two main tracks is approximately 13 percentage points. The true effect may be even smaller, as our observational data cannot entirely eliminate residual selection effects. Given these limitations and the outcomes of our robustness checks, the 13-percentage-point estimate can be interpreted as an upper bound of the effect size for highly motivated students. These findings are both positive and encouraging. They suggest that final educational success in secondary education is not strongly dependent on the initial track placement, provided that students and their parents are highly motivated to attain a high-level qualification. In many cases, this goal is achieved even when students begin their secondary education in non-academic tracks. It is also notable that, for these students, academic achievement and satisfaction are not negatively impacted by placement in non-academic tracks. In summary, diversion effects for positively selected, highly motivated students are likely smaller than previously assumed.

However, we would also like to emphasize that even this rather small difference means that some students cannot fulfil their full potential and end up with educational degrees that are below what they could possibly achieve. This loss is relevant and should be considered in future educational reforms. Policy makers should be aware of the differences between the academic and the non-academic tracks and attempt to make the system even more transparent and permeable. The good news is that the differences between the track types are apparently not due to differences in ability development, which means that teaching plans or teacher education, which are most relevant for the development of competencies, are not the main source of inequality. One could assume that differences in information and navigating the (rather complex) system, especially after the completion of a first degree, are more important. Students and parents should be made aware of the fact that good educational performance allows educational upgrading and completing the highest degree is always possible, even if students start their trajectories on a non-academic track. Making this information more salient and supporting especially the high-performing students in their journey through the educational system might be a possibility to further reduce the inequality between the tracks. We suggest to conduct further empirical investigations to check whether this aspect of navigating the system and information is relevant as this study can only provide first suggestions that require empirical validation. One quite different aspect that should not be overlooked is the fact that entering the Dual system and learning a trade, instead of entering tertiary education, is a completely valid choice. Students of all school tracks have the option to make this choice, which is even vital to society as a whole as craftspeople will always be required. In this vein, future analyses might want to investigate decision-making processes of secondary school students in more detail, potentially using qualitative methods.

To conclude, we recommend further analysis of the factors that contribute to the emergence of diversion effects, as this would enhance our understanding of when and why such effects occur. Gaining this insight could support the development of targeted strategies and policies aimed at mitigating the potential negative consequences of diversion in the long term. Finally, adopting a broader perspective on the implications of diversion effects for the overall quality and equity of education in Germany represents an important and promising direction for future research.

## CRediT authorship contribution statement

**Felix Bittmann:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Ethical approval

The NEPS study is conducted under the supervision of the German Federal Commissioner for Data Protection and Freedom of Information (BfDI) and in coordination with the German Standing Conference of the Ministers of Education and Cultural Affairs (KMK) and – in the case of surveys at schools – the Educational Ministries of the respective Federal States. All data collection procedures, instruments and documents were checked by the data protection unit of the Leibniz Institute for Educational Trajectories (LIfBi). The necessary steps are taken to protect participants' confidentiality according to national and international regulations of data security. Participation in the NEPS study is voluntary and based on the informed consent of participants. This consent to participate in the NEPS study can be revoked at any time.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of Competing Interest

The authors declare that they have no conflict of interest.

## Acknowledgements

Special thanks to Marie-Louise Nilles for her support in reviewing the literature. Additional thanks to the journal editor and the multiple reviewers for their helpful and constructive feedback.

## Appendix

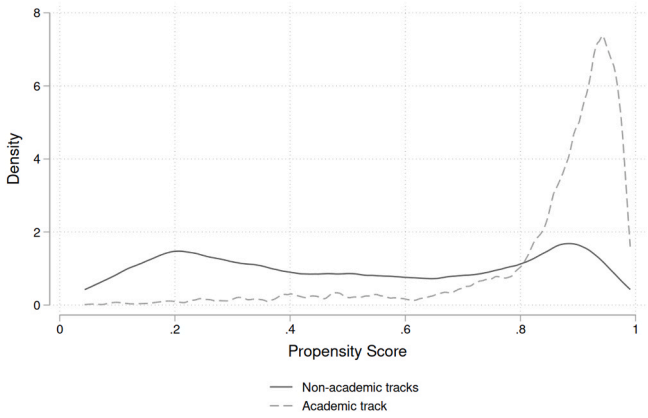


**Figure A1.** Propensity scores after having selected the region of common support for both the treatment and the control group. The region of common support has been using the overlap criterion as implemented in *kmatch*

## Robustness checks

In this appendix we conduct various robustness checks to see how stable the main findings are. Three additional analyses or versions of the initial analyses are conducted.

First robustness check: the binding character of the teacher recommendation in primary schooling is the most relevant structural factor as it limits the parental choice for a track. We have accounted for this problem by including this information as a control variable in the main analyses, yet this might not be enough to enable a fair comparison. As a further check, we removed all students in federal states with a binding recommendation from the analyses. While this reduces the sample size to 1065, testing how strongly the conclusions differ from the main model is sensible. We start with the longitudinal analyses and present the final results in Table A1. Again, the interaction term between wave and track is of interest. As we can see, no coefficient is statistically significant at the 5 % level. This indicates that achievement and satisfaction develop similarly in both tracks.

Regarding the outcome of whether the *Abitur* has been obtained or not, we present the results in Table A2. Here, the numbers are also highly similar to the main results. While the unadjusted model reports large differences between the two tracks, the two other models agree on a large reduction of the effect as soon as relevant controls are considered. Summarized, excluding students in federal states where the binding teacher recommendation was still active does not influence our conclusions.

Second robustness check: so far, we have focussed on idealistic educational aspirations, which only express wishes and desires but do not consider any limitations. These can be severe, and insufficient academic performance is a large obstacle to obtaining high educational qualifications in the German system, which is designed to select on ability. How do the conclusions change if one applies the same selection criterion for realistic educational aspirations? To answer this question, we repeat the main analyses.[9] This results in an even stricter selection since realistic aspirations imply not only a high motivation to obtain the degree but also sufficient academic performance and other resources to complete the degree. Note that the sample size is reduced to 1348 by this selection (15.1 % in the non-academic tracks / 84.9 % in the academic track). Regarding the development of achievement and satisfaction (Table A3), we also see no differences between the academic track and other tracks, just as in the main analyses. Concerning the final degree obtained, we see that the differences are smaller overall between the two models (Table A4). This also makes sense since by enforcing an overall stricter sample, which is even more positively selected from the beginning, the influences of (self-) selection should be weaker. However, accounting for selection effects still reduces the gaps between the two groups.

To summarize our robustness checks so far, the main results are stable and hold up well. No single model specification shows that the two groups' trajectories of achievement or satisfaction differ. Regarding the highest educational qualification obtained in secondary education, we see that students' advantage in the academic track is always largely reduced by accounting for selection effects. The residual differences are rather small. This conforms with the main analyses and demonstrates that the diversion effect of non-academic tracks is rather small.

Third robustness check: We would like to analyse the potential influence of unobserved confounders. Any study that uses observational data can suffer from omitted variable bias, and no dataset can include all potential confounders. This is a severe issue, as violating this assumption can lead to biased and misleading results. To combat this problem, various approaches are available to assess how strongly the results would change if unmeasured confounders were present. In this check, we follow the approaches suggested by Oster (2019) and Diegert et al. (2022), the latter being an extension of the Oster approach. The idea behind these analyses is as follows: while we can easily compute regression coefficients on the observed data, how stable is this coefficient if we assume that unmeasured confounders are present? The classical approach by Oster is to check how large the influence of unobserved variables needs to be to drive the regression coefficient to zero ("explain it away"). This factor (observed to unobserved influences) is termed Oster's Delta. We use the Stata package *psacalc* to compute it. Under the assumption that $R_{max}$ is 100 %, which means that in the model with all observed and unobserved variables, 100 % of the variation in the outcome variable can be explained, Delta is 0.096. However, in the social sciences, such a high $R_{max}$ value is rather unrealistic as even the best models can only explain an incomplete share in the total outcome variation, since human life is highly variable and arbitrary. If we assume that $R_{max}$ is 50 %, Delta is 0.250; if we assume it is 30 %, it is 0.695. Keep in mind that the factual explained variance (compare Table 3) is 19.1 %, so the influence of unobservables would need to be immense to reach that high levels of variance explanation.

The approach by Diegert et al. is a bit more flexible as it allows confounders to be correlated with observed controls, which is a reasonable assumption in the social sciences. We utilize the Stata package *regsensitivity* to compute the bounds for the binary treatment variable "Track" in the regression analysis with the same control variables as reported before. As stated, the coefficient is 0.115 (see Table 3). The question is, how would this

[9] The filial realistic item is as follows: "Considering everything you know now: What qualification will you actually leave school with?"

coefficient change in the presence of unobserved confounders? Usually, researchers are interested in the breakdown point, which gives the point at which the unobserved confounders are so strong that the sign of the coefficient switches. This is depicted graphically in Figure A2.

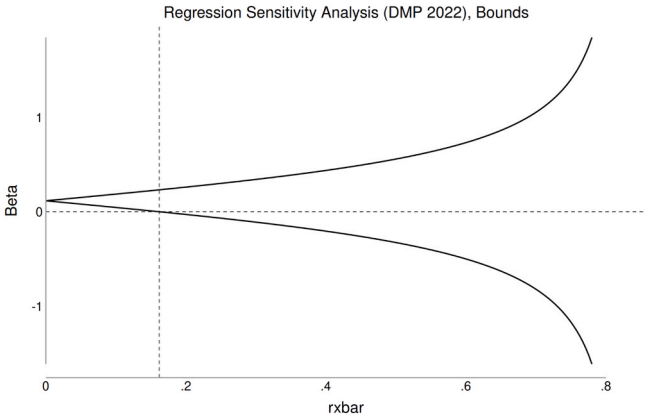


**Figure A2.** Change in the coefficient of track in the presence of unobserved confounders

Source: NEPS SC3, imputed data (M = 40). The vertical dashed line indicates the breakdown point where the regression coefficient of "track" would switch its sign and becomes negative. Rxbar = 0 is the factual reported regression coefficient.

The breakdown point is reached when *rxbar* reaches 16.2 %.[10] This means, if the unobserved influence is about 16 % of the strength of the observed variables, the regression coefficient of school track becomes zero. Remember: While we can never rule out the presence of unobserved confounders, we are rather confident that our dataset contains a rich set of relevant controls backed up by theoretical arguments and previous research results. In any case, researchers should be aware of the possibility of bias; this analysis helps gauge the magnitude of this potential bias. Further checks, such as varying the correlation between observed and unobserved controls, show that the breakdown point is virtually unaffected by doing so. We have also applied the same sensitivity check to both robustness versions discussed above and obtained lower upper bounds (22 % for the federal state check and 17 % when using realistic instead of idealistic aspirations).

**Robustness check: only non-binding federal states**

**Table A1**
Numerical results for the growth-curve models

| | Math achievement | Reading achievement | Life satisfaction | School satisfaction |
|---|---|---|---|---|
| Academic track | 0.106 | 0.128 | 0.0355 | 0.0832 |
| | [ 0.156,0.369] | [ 0.457,0.201] | [ 0.469,0.540] | [ 0.716,0.550] |
| Wave | 0.173*** | 0.254*** | 0.0563* | 0.0846** |
| | [0.149,0.197] | [0.225,0.282] | [ 0.103, 0.00959] | [ 0.142, 0.0276] |
| Academic track X Wave | 0.00320 | 0.00954 | 0.0168 | 0.0189 |
| | [ 0.0320,0.0256] | [ 0.0252,0.0443] | [ 0.0713,0.0376] | [ 0.0863,0.0484] |
| Constant | 0.374** | 0.762*** | 8.344*** | 8.097*** |
| | [ 0.600, 0.148] | [ 1.043, 0.482] | [7.909,8.779] | [7.555,8.640] |
| Individuals | 1065 | 1065 | 1065 | 1065 |
| Observations | 4260 | 4260 | 8520 | 8520 |

**Table A2**
Outcomes regarding achieving higher education eligibility

| | Raw (Unadjusted) | Matching | Regression |
|---|---|---|---|
| Transition to the academic track | 0.234***[0.166; 0.303] | 0.086*[0.005; 0.168] | 0.104**[0.033; 0.174] |
| Adj. $R^2$ | 0.081 | - | 0.212 |
| Observations | 1065 | 1065 | 1065 |

[10] The concrete meaning of *rxbar* is slightly technical and defined as "…a unit-free measure of the relative magnitude of selection on unobservables to selection on observables. [rxbar] = 0 corresponds to the baseline assumption ($\pi 2 = 0$) [no unobserved confounders present] while [rxbar] > 0 allows for some selection on unobservables (Diegert et al., 2022: 10 f.).

## Robustness check: selection on realistic aspirations (instead of idealistic ones)

**Table A3**
Numerical results for the growth-curve models

| | Math achievement | Reading achievement | Life satisfaction | School satisfaction |
|---|---|---|---|---|
| Academic track | 0.0369 | 0.0425 | 0.0840 | 0.124 |
| | [ 0.369,0.295] | [ 0.380,0.465] | [ 0.498,0.666] | [ 0.738,0.490] |
| Wave | 0.159*** | 0.247*** | 0.0704* | 0.118*** |
| | [0.123,0.194] | [0.207,0.288] | [ 0.127, 0.0135] | [ 0.179, 0.0565] |
| Academic track X Wave | 0.0151 | 0.00885 | 0.0160 | 0.0255 |
| | [ 0.0211,0.0513] | [ 0.0507,0.0330] | [ 0.0760,0.0439] | [ 0.0912,0.0402] |
| Constant | 0.0278 | 0.522* | 8.534*** | 8.659*** |
| | [ 0.357,0.302] | [ 0.936, 0.108] | [7.979,9.088] | [8.086,9.232] |
| Individuals | 1348 | 1348 | 1348 | 1348 |
| Observations | 5392 | 5392 | 10784 | 10784 |

**Table A4**
Outcomes regarding achieving higher education eligibility

| | Raw (Unadjusted) | Matching | Regression |
|---|---|---|---|
| Transition to the academic track | 0.152*** | 0.104* | 0.083* |
| | [0.094; 0.210] | [0.004; 0.204] | [0.012; 0.154] |
| Adj. $R^2$ | 0.026 | - | 0.110 |
| Observations | 1348 | 1348 | 1348 |

## Trimming of entropy balancing weights

By removing extreme entropy balancing weights, it can be tested whether the main diversion results are influenced by special cases with very high weights. To achieve this, these extreme cases are systematically discarded and results are compared to the main model. We test three different models: removing weights larger than 20 (affecting 15 cases), larger than 10 (affecting 50 cases), and larger than 6.1 (affecting 95 cases or 5 % of the total sample). While 5 % of the total sample does not seem to be much, keep in mind that the track variable is very unbalanced as there are, in total, only 429 students in the sample which attend non-academic tracks (see Table 1). This 5 % rule removes about 21 % of the non-academic students and is, hence, a rather severe reduction of power and information. The results are shown in Figure A3.

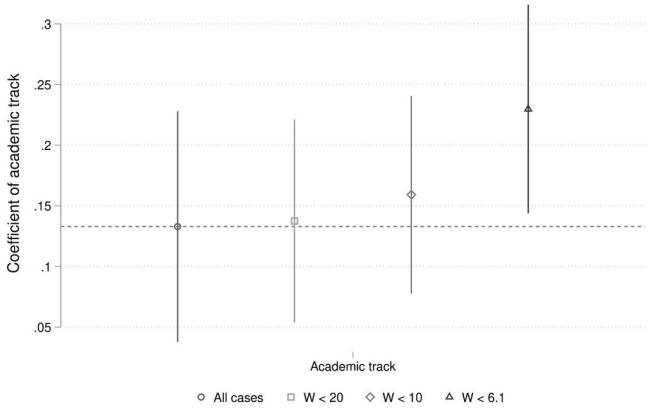


**Figure A3.** Change in the coefficient of track when removing cases with extreme balancing weights. Scenario "W <6.1" corresponds to removing the top 5 % of largest weights. The dashed line reports the original EB coefficient of track.

As shown, the estimated diversion effect tends to increase as more cases are removed from the analysis. While this suggests that the main model may slightly underestimate the diversion effect, several considerations should be kept in mind when interpreting this robustness check. First, selectively removing cases can steer any analysis in a particular direction and should therefore be undertaken with caution. The fact that a more restrictive model yields a stronger effect does not necessarily mean it is more accurate or superior to the main model, as it is based on a different, narrower sample of students. It merely illustrates one possible scenario. Second, the OLS regression reported in Table 3 yields a coefficient very similar to that produced by the entropy balancing (EB) approach (0.115). This consistency across methods indicates that the results are not solely driven by the choice of EB, but are instead robust to alternative, widely used estimation strategies. Both methods aim to approximate a counterfactual scenario in which students from different tracks are comparable on all observed covariates. Of course, this does not reflect actual conditions but instead simulates a "what-if" situation. Third and finally, it is important to note that all models reported in this study are restricted to the region of common support. This ensures that the comparisons are based on overlapping groups of students and prevents the formation of extreme, non-comparable cases, thereby supporting the validity of the findings.

## Data Availability

Data are available after registration as a researcher from https://doi.org/10.5157/NEPS:SC3:12.1.0. Stata Do-files are available from https://github.com/fbittmann/replication_see.

## References


Allison, P. D. (2001). Sage publications. *Missing data, 136*.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Bayer, S. (2020). *Mathematikunterricht im Vergleich zwischen den Schularten. In Empirische Erziehungswissenschaft 76, Eds. Becker, Rolf, Sigrid Blömeke, Wilfried Bos, Hartmut Ditton, Cornelia Gräsel, Eckhard Klieme, Kai Maaz, Thomas Rauschenbach, Hans-Günther Roßbach, Knut Schwippert, Ludwig Stecher, Christian Tarnai, Rudolf Tippelt, Rainer Watermann and Horst Weishaupt. Münster: Waxmann*. 10.31244/9783830992479.

Baumert, J., Stanat, P., & Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern-und Entwicklungsmilieus. In *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit: Vertiefende Analysen im Rahmen von PISA 2000* (pp. 95–188). Wiesbaden: VS Verlag für Sozialwissenschaften.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104*(3), 682.

Becker, R., & Hecken, A. E. (2008). Warum werden Arbeiterkinder vom Studium an Universitäten abgelenkt? Eine empirische Überprüfung der „Ablenkungsthese" von Müller und Pollak (2007) und ihrer Erweiterung durch Hillmert und Jacob (2003). *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 60*(1), 7–33. https://doi.org/10.1007/s11577-008-0001-1

Bittmann, F. (2021). Academic track mismatch and the temporal development of well-being and competences in German secondary education. *Vienna Yearbook of Population Research, 19*, 1–36. https://doi.org/10.1553/populationyearbook2021.res5.1

Bittmann, F. (2023). High hopes but aiming low? Explaining why some families with high educational aspirations do not choose the academic track in German secondary education. *SN Social Sciences, 3*(11), 194. https://doi.org/10.1007/s43545-023-00780-x

Bittmann, F., & Schindler, S. (2021). Analysing diversion processes in german secondary education: School-track effects on educational aspirations. *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 73*(2), 231–257. https://doi.org/10.1007/s11577-021-00789-1

Blossfeld, H., & Roßbach, H. (2019). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (Second ed.). VS Verlag für Sozialwissenschaften. ZfE, Vol. 3.

Diegert, P., Masten, M. A., & Poirier, A. (2022). Assessing Omitted Variable Bias when the Controls are Endogenous (Version 4). *arXiv*. https://doi.org/10.48550/ARXIV.2206.02303

Dudek, P. (1993). *Transformationen der deutschen Bildungslandschaft: Lernprozeß mit ungewissem Ausgang* (Vol. 30). Beltz.

Eckhardt, T. (2017). *The Education System in the Federal Republic of Germany 2015/2016. A description of the responsibilities, structures and developments in education policy for the exchange of information in Europe.* ⟨https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-engl-pdfs/dossier_en_ebook.pdf⟩.

Esser, H., & Seuring, J. (2020). Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit: Theoretische Modellierung und empirische Analyse der Effekte einer strikten Differenzierung nach den kognitiven Fähigkeiten auf die Leistungen in der Sekundarstufe und den Einfluss der sozialen Herkunft in den deutschen Bundesländern mit den Daten der „National Educational Panel Study "(NEPS). *Zeitschrift für Soziologie, 49*(5-6), 277–301.

Esser, H. (2024). Kein Vergleich! Zur „kritischen Betrachtung des Model of Ability-Tracking (MoAbiT)" von Lorenz, Lenz und Rjosk (2023) in der Zeitschrift für Soziologie. *Zeitschrift für Soziologie, 53*(4), 419–425. https://doi.org/10.1515/zfsoz-2024-2024

Fischer, L., Rohm, T., Gnambs, T., & Carstensen, C. H. (2016). Linking the data of the competence tests. *NEPS Survey Paper, 1*. ⟨https://www.lifbi.de/Portals/2/Survey%20Papers/SP_I.pdf⟩.

Gerth, M. (2022). Is what they aspire what they get? The role of cultural heterogeneity in the classroom for the realization of Abitur aspirations. https://doi.org/10.31235/osf.io/tynpf.

Gnambs, T., & Lockl, K. (2023). Bidirectional effects between reading and mathematics development across secondary school. *Zeitschrift Für Erziehungswissenschaft, 26*(2), 345–371. https://doi.org/10.1007/s11618-022-01108-w

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis, 20*(1), 25–46. https://doi.org/10.1093/pan/mpr025

Hartlaub, V., & Schneider, T. (2012). *Educational choice and risk aversion: How important is structural vs. Individual risk aversion? SOEPpapers on multidisciplinary panel data research*. ⟨https://www.econstor.eu/bitstream/10419/59027/1/717167135.pdf⟩.

Hartung, A., & Weßling, K. (2024). Discouraged and hedged – why students enter VET after obtaining university eligibility. *Journal of Vocational Education Training, 0*(0), 1–24. https://doi.org/10.1080/13636820.2024.2307536

Hascher, T., & Hagenauer, G. (2011). Schulisches Wohlbefinden im Jugendalter–Verläufe und Einflussfaktoren. *Jahrbuch Jugendforschung*, 15–45. https://doi.org/10.1007/978-3-531-93116-6_1

Henninges, M., Traini, C., & Kleinert, C. (2019). Tracking and sorting in the German educational system. *Leibniz Institute for Educational Trajectories (LIfBi) Working Paper No, 83*. ⟨https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXXXIII.pdf⟩.

Herke, M., Rathmann, K., & Richter, M. (2019). Trajectories of students' well-being in secondary education in Germany and differences by social background. *European Journal of Public Health, 29*(5), 960–965. https://doi.org/10.1093/eurpub/ckz049

Hillmert, S., & 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. https://doi.org/10.1016/j.rssm.2009.12.006

Holtmann, M., Rjosk, C., Jansen, M., & Lüdtke, O. (2024). The double-edged sword of de-tracking reforms in Germany: Using a Difference-in-Differences approach to estimate effects on students' educational aspirations and their association with students' socioeconomic backgrounds. *Studies in Educational Evaluation, 81*, Article 101350. https://doi.org/10.1016/j.stueduc.2024.101350

Hurrelmann, K. (2013). Das Schulsystem in Deutschland: Das" Zwei-Wege-Modell" setzt sich durch. *Zeitschrift Für Pädagogik, 59*(4), 455–468.

Jann, B. (2017). *KMATCH: Stata module module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment*, [Stata]. ⟨https://ideas.repec.org/c/boc/bocode/s458346.html⟩.

Karlson, K. B. (2015). Expectations on Track? High School Tracking and Adolescent Educational Expectations. *Social Forces, 94*(1), 115–141. https://doi.org/10.1093/sf/sov006

Karlson, K. B. (2019). College as equalizer? Testing the selectivity hypothesis. *Social Science Research, 80*, 216–229. https://doi.org/10.1016/j.ssresearch.2018.12.001

Klein, D. (2014). *MIMRGNS: Stata module to run margins after mi estimate* [Stata]⟨https://ideas.repec.org/c/boc/bocode/s457795.html⟩.

Kristen, C. (2002). Hauptschule, Realschule oder Gymnasium? *KZfSS Kölner Zeitschrift Für Soziologie Und Sozialpsychologie, 54*(3), 534–552. https://doi.org/10.1007/s11577-002-0073-2

Lorenz, G., Lenz, S., & Rjosk, C. (2023). Effizienz und soziale Ungleichheit in strikt leistungsdifferenzierenden Bildungssystemen. Eine kritische Betrachtung des Model of Ability Tracking (MoAbiT). *Zeitschrift für Soziologie, 52*(4), 404–424.

Müller, W., & Pollak, R. (2004). Weshalb gibt es so wenige Arbeiterkinder in Deutschlands Universitäten? In R. Becker, & W. Lauterbach (Eds.), *Bildung als Privileg? Erklärungen und Befunde zu den Ursachen der Bildungsungleichheit* (pp. 311–352). https://doi.org/10.1007/978-3-322-93532-8_11

Pohl, S., & Carstensen, C.H. (2013). *Scaling of competence tests in the National Educational Panel Study—Many questions, some answers, and further challenges*. https://doi.org/10.25656/01:8430.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics, 37*(2), 187–204.

Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology, 82*(4), 647–671.

Schindler, S. (2014). *Wege zur Studienberechtigung – Wege ins Studium?: Eine Analyse sozialer Inklusions- und Ablenkungsprozesse. VS Verlag für* Sozialwissenschaften. //⟨www.springer.com/de/book/9783658038403⟩.

Schindler, S., & Bittmann, F. (2021). Diversion or Inclusion? Alternative Routes to Higher Education Eligibility and Inequality in Educational Attainment in Germany. *European Sociological Review, 37*(6), 972–986. https://doi.org/10.1093/esr/jcab025

Sewell, W. H., Hauser, R. M., Springer, K. W., & Hauser, T. S. (2003). As we age: A review of the Wisconsin Longitudinal Study, 1957–2001. *Research in Social stratification and Mobility, 20*, 3–111. https://doi.org/10.1016/S0276-5624(03)20001-9

Traini, C., Kleinert, C., & Bittmann, F. (2021). How does exposure to a different school track influence learning progress? Explaining scissor effects by track in Germany. *Research in Social Stratification and Mobility*. , Article 100625. https://doi.org/10.1016/j.rssm.2021.100625

Van den Broeck, L., Demanet, J., & Van Houtte, M. (2020). The forgotten role of teachers in students' educational aspirations. School composition effects and the buffering capacity of teachers' expectations culture. *Teaching and Teacher Education, 90*, 103015. https://doi.org/10.1016/j.tate.2020.103015

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (Second ed.). MIT Press.

Wu, Y. J., & Becker, M. (2023). Association between school contexts and the development of subjective well-being during adolescence: A context-sensitive longitudinal study of life satisfaction and school satisfaction. *Journal of Youth and Adolescence, 52*(5), 1039–1057. https://doi.org/10.1007/s10964-022-01727-w