More Teachers, Smarter Students? Potential Side Effects of the German Educational Expansion*

Matthias Westphal
Paderborn University,
RWI Essen
&
Ruhr Graduate School in Economics

November 2017

Abstract

In this paper, I evaluate potential side effects of the educational expansion in Germany on the learning outcomes of today's students. The educational expansion was a demand shock in the labor market of teachers, which could have thus encouraged individuals with different teaching abilities to eventually become teachers. I find that replacing a non-affected teacher with an educational expansion teacher leads to a 2 percent reduction in students' test scores. Explorative analyses suggests that these teachers are more extrinsically rather than intrinsically motivated. The results highlight that monitoring and investing in quality is important for future extensions of public institutions.

Keywords: Human capital acquisition, teacher effectiveness, educational expansion
JEL Classification: *H75, I20, I21, I28, J24*

Matthias Westphal: Paderborn University, Warburger Straße 100, 33098 Paderborn, Germany, E-mail: matthias.westphal@upb.de.

*I am grateful to Hendrik Schmitz for many valuable comments, in particular those related to the improvement of the framing of the research question. Also I would thank Markus Nagler for his precious feedback that improved the paper significantly. Furthermore, I thank Christian Bünnings, Martin Fischer, Ilka Gerhardts, Matthias Giesecke, Daniel Kamhöfer, Irene Palnau, Valentin Schiele, Anna Werbeck, participants of the health economics research seminar of the University of Duisburg-Essen, the TEAM seminar in Paderborn, the 10th RGS Doctoral Conference in Economics, the EALE Conference in St. Gallen and the therapy seminar of the RWI Essen. Financial support by the Fritz-Thyssen Stiftung (grant no. 20.16.0.069) is gratefully acknowledged.

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohorts Class 5, doi:10.5157/NEPS:SC3:5.0.0 and Class 9 doi:10.5157/NEPS:SC4:7.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

1 Introduction

In recent years, the view that has ultimately prevailed is that education throughout the life course is important for acquiring skills that are decisive for, but not exclusively confined to, the labor market (Heckman et al., 2010; Chetty et al., 2011; Zimmerman, 2014; Kamhöfer et al., 2017). Teachers have a key role in creating environments and incentives for students to acquire these important skills, typically referred to in economics as the acquisition of human capital (Hanushek, 1971; Hanushek and Rivkin, 2006; Chetty et al., 2014a). Because of this key role, it is important to look at the leverage of educational policy on attracting high-quality teachers. If, for example, relatively less suitable individuals take up the teaching profession in response to changes of institutional arrangements, they could have a negative impact on the performance of their students. As all teachers educate generations of pupils over the course of their career, teachers can have a highly persistent impact on the skill acquisition of these pupils. Evidence from recent studies advocates such a persistent impact of teachers since resulting skill differentials at school may well spill over to later life by, for instance, affecting labor market performance (Chetty et al., 2014b).

In Germany, as in most industrialized societies, in the second half of the past century, educational policies were at the core of government institutional reforms. The goal was to increase access in particular to higher secondary education, namely the intermediate track (Realschule) and the academic track (Gymnasium), relative to the then-dominant basic track.[1] The quantitative expansion of both tracks was substantial even in relative terms: whereas only 20 percent of all pupils went to either one of both tracks in the 1960s, this share had doubled by the end of the 1980s. This tremendous increase led to an upsurge in the demand for teachers.[2] Due to the educational expansion, roughly 150,000 new positions as teachers were created. These positions could not even theoretically be filled with basic track teachers, as these positions required more formal training.[3]

Did the implementation of this quantitative expansion lead to a diminishing quality of teachers? If, at any time, only the most motivated and able individuals took up the profession, an unanticipated and unprecedented increase in the demand for teachers could have encouraged less motivated and able individuals to eventually become teachers. The educational expansion is not only important because it created a demand-side variation in the labor market for teachers, it also captures a highly policy-relevant effect. Many of

[1] At the same time, comprehensive schools (Gesamtschulen) were introduced. This school track, however, only played a minor role.

[2] Because of a coinciding reduced student-teacher ratio, the demand for teachers was even higher than the increase in student numbers.

[3] In addition, the overall number of students in secondary education mechanically increased due to the changing track composition (academic track required four more years of schooling; the intermediate track, one year).

today's policies are often targeted at expanding public institutions like, for instance, the recent extension of the daycare sector and – potentially – of the future formal long-term care sector in Germany. These expansions exhibit characteristics that are similar to the educational expansion in the 1970s and '80s. Hence, knowledge about the past expansion is informative about how to efficiently implement new ones in the future.

The literature on teacher selection and its effects on student performance initially focused on identifying determinants of teacher selection. There is a large strand of literature that looks at the role of wage differentials between teachers' and the outside labor market (see, for instance, Britton and Propper, 2016, Loeb and Page, 2000, and Figlio, 1997, among others). Nagler et al. (2015) examine the consequences of business cycle-induced teacher selection on students' test scores. These studies find that a larger wage-differential leads to a diminishing teacher quality. Beyond wages, there are also further characteristics of the labor market of teachers subject to some studies. For instance, Lakdawalla (2001) determines the role of technological change and Bacolod (2007) considers the soared acceptance of female teachers. These studies likewise detect that teachers react to changed external incentives. Chetty et al. (2014b) go one step back by identifying the general impact of teachers on the human capital acquisition of their students. They uncover that replacing an average teacher with a teacher from the 5% quantile of the distribution of teacher quality raises the net present value of their lifetime earnings of the affected students by $250,000 per classroom.

I contribute to the literature on teacher selection and its effects on student performance mainly in two ways. First, this is the first study to specifically assess the consequences of one particular and major social change of the last 60 years – the educational expansion – not on those who are taught[4] but rather on those who teach. Insights into teachers are important since they are under a more direct control of policymakers who could then apply these insights to modifying the hiring process of teachers. Second, I am able to provide evidence on a much more homogeneous group of high-skilled pupils who attend the academic track in Germany. This is in contrast to the existing literature that looked primarily at the comprehensive school system of the US or the UK.

To substantiate the exact specification of the educational expansion rate and the subsequent interpretation of the effects, I employ a simple theoretical framework of how marginal teachers affect the average quality of all teachers of a certain cohort. This model corroborates using relative changes in the stock of teachers in the federal state and year of the high school graduation as the educational expansion rate. This rate proxies the conditions of the teachers' labor market (and coinciding career incentives for those who are encouraged to become teachers). Subsequently, this proxy is related to the test scores

[4]Studies that focus on students comprise Siegler (2012) and Kamhöfer et al. (2017) for tertiary education, as well as Jürges et al. (2011) for secondary education.

of students instructed by a teacher decades later. By using these changes within German federal states that control and legislate the educational system within their borders, I am able to isolate the overall effects from a wide range of other effects. These confounding effects may arise because of unobserved third factors, for example, effects that go along with teachers' general experience or, more importantly, potential persistent differences in the quality of the educational system of the federal state. Concerning the former, for instance, the students' performance is measured decades later, long after the educational expansion was complete. Hence, I can disentangle the effects of the educational expansion that operates through teachers from the repercussions on students. Furthermore, I use a between-subjects difference-in-differences model to address the concern that good teachers may want to teach at good schools with better students. In the absence of any spillover effects, estimates of the cross-subject teacher environment on student test score relations (math teacher, reading scores and German teacher, math scores) identify confounding school selection effects, which can then be differenced out from the same-subject effects. If this teacher skill differential of educational expansion teachers is indeed driving the effect, I would expect this skill differential to also be reflected in some observed characteristics, such as subjectively assessed measures on intrinsic and extrinsic motivation.

To summarize the results, I find that students taught by teachers who witnessed an expanding teacher force in their federal state just after high school graduation score less in math and reading competence tests. By decomposing the effect into a component that is due to school selection (correlation between good teachers and initially good students) and a direct effect on test scores, I find that a significant share of the overall effect can be attributed to the direct effect of teachers on students. Teachers who graduated from high school in an average expansion year reduce the test scores of their students by 2 percent of an unconditional standard deviation (sd) relative to teachers that graduated in years with no expansion. The magnitude of the effect is comparable to related studies and is non-negligible. In providing an explanation for the identified test score differential, I look at the reported grade of the teachers' high school exit exam (Abitur) and examine further subjective measures of job choice and work ethic. I find that the educational expansion rate weakly predicts the academic achievement of teachers. In addition, educational expansion teachers are more extrinsically rather than intrinsically motivated.

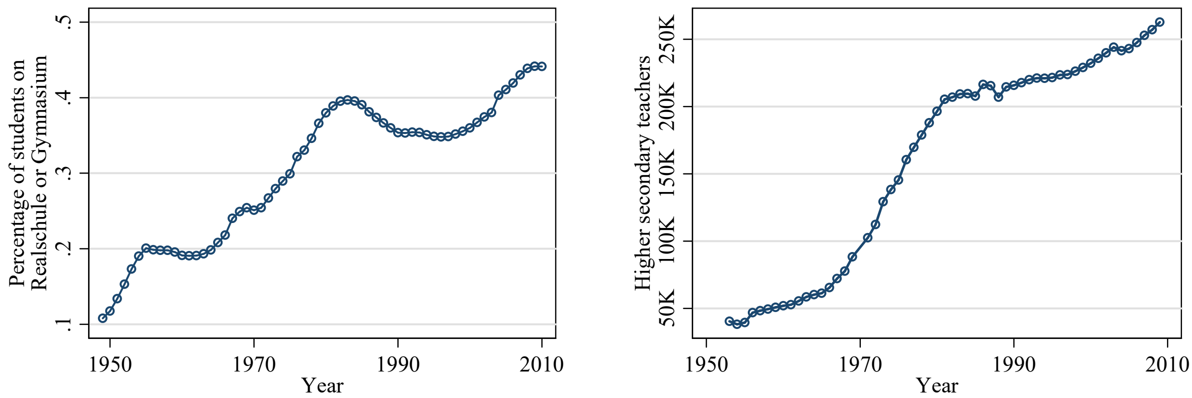
The results have at least two important implications. First, as the policymaker certainly has more leverage in hiring good teachers than on directly influencing students or their family background, the conclusions of this paper are important for shaping future policies. Connected to this, the second implication concerns today's and future expansions of public institutions in general, which become increasingly necessary in changing societies. Given the results of this paper, it seems crucial to not only invest in quantitative aspects, such as increasing the scope of arguably beneficial public institutions. Qualitative aspects are an important margin to invest in when implementing the expansion of these institutions. The substantial ongoing extension of daycare facilities (day nurseries and preschools) serves as a prime example. Since the educational expansion is paralleled by this expansion of daycare facilities, the results of this paper can rather easily be extrapolated to this setting.

The remainder of the paper is structured as follows: Section 2 sets out the institutional background of the educational expansion in general and the teacher market in particular. Section 3 presents the empirical strategy that aims at estimating causal effects. Subsequently, a small theoretical mechanism is introduced that justifies the specification of the educational expansion rate and facilitates the interpretation of the results. Section 4 presents the data. Section 5 shows the main results of students' learning outcomes, assesses its robustness and presents supporting evidence on the characteristics of educational expansion teachers. Finally, Section 6 concludes.

2 The educational expansion and the market for teachers in Germany

In Germany, at least three things changed the notion of the scope of higher education, all of which took place roughly within 15 years. First, the view ultimately prevailed that education was key for social participation as a citizen, which served as a powerful intellectual and publicly influential argument to promote education (Dahrendorf, 1965). Second, as a consequence of its increased role internationally, reports of the OECD showed that Germany's system was internationally underdeveloped. This had, not least because of an influential book (Picht, 1965, which based on arguments set out in Picht, 1964), a huge impact on public opinion. The new and changed notion of education was reflected by the Social Democratic Party (SPD) making it the cornerstone of their new programmatic orientation: education policy was granted federal political importance by a party whose clientele traditionally came from educationally deprived strata (Osterroth and Schuster, 2000). Third, because of the Sputnik crisis in 1957, Western societies realized that they were trailing behind the Soviet Union. Opening higher secondary and tertiary education for a broader population was identified as being important for closing this gap in the long run. All these developments led to changes mainly in the supply of education, which shifted the composition of the students in terms of their field of study from public institutions traditionally being the most important employer of university graduates toward newly created jobs in engineering, administration, and the business sector (see, for example, Lundgreen and Schallmann, 2013).

The educational expansion also substantially affected secondary schools. This is visible in Figure 1a, where the share of pupils in the intermediate and academic track is plotted over time. The increased number of pupils required more teachers, also because shifts in the track composition led to a mechanical increase in the average years of schooling (the intermediate track had one more year of schooling, the academic track four years more). Figure 1b illustrates the upsurge in teacher positions in higher secondary education over time: within 20 years, 150,000 additional teacher positions were created. The long-term repercussions of these new teachers are the subject of this paper. This requires looking at the dynamics that took place simultaneously, concerning, among others, teacher remuneration and the education of teachers in Germany. The current process of teacher



(a) The percentage of intermediate and academic track students over time

(b) Dynamics in the stock of teachers on intermediate and academic track schools

Figure 1: Impact of the German educational expansion

Source: Köhler and Lundgreen (2015)

training in Germany was implemented in 1917 for academic track teachers and was extended to include all teachers at primary and secondary schools up until 1970 (Köhler and Lundgreen, 2015). This process is called the "academization of the teaching profession" (Bölling, 1983; Köhler and Lundgreen, 2015). The training of all teachers from at least 1970 onward is set up as a two-stage process. All high school graduates with an academic track education (Abitur) are in principle eligible to be trained as teachers. Initially, teachers are educated at a university, commonly graduate in two specific subjects (Erstes Staatsexamen) and start a more education-specific vocational training at a certain school. After graduation from university – which takes 4.5 years – teachers graduate a second time (Zweites Staatsexamen) – which takes an additional 1-2 years – where teaching skills are tested. At the same time, there were also some changes in how teachers were remunerated. For example, one consequence of the academization was an increase in the salary level of teachers (Bölling, 1983). In addition, the teacher salary was leveled up to reduce the excess demand of teachers and to match their salary to wages in professions that required a similar qualification level. This, however, was largely completed before

1970 (Bölling, 1983; Köhler and Lundgreen, 2015) and therefore does not interfere with the study period (from 1970 onward).

3 Empirical strategy and theoretical mechanism

3.1 Empirical strategy

The aim is to compare "educational expansion" teachers (EETs) with teachers who were not influenced by the educational expansion. I consider EETs as being individuals who started their teacher training and education during the massive demand increase that occurred during the educational expansion. On average, these teachers may differ because of some marginal teachers. These marginal teachers are a subset of all EETs and only took up the teaching profession because of changed career incentives (Ashraf et al., 2014). For instance, an awareness of the possibility of eventually becoming a teacher may have surged. If the educational expansion occurred in certain years and not in others, I could

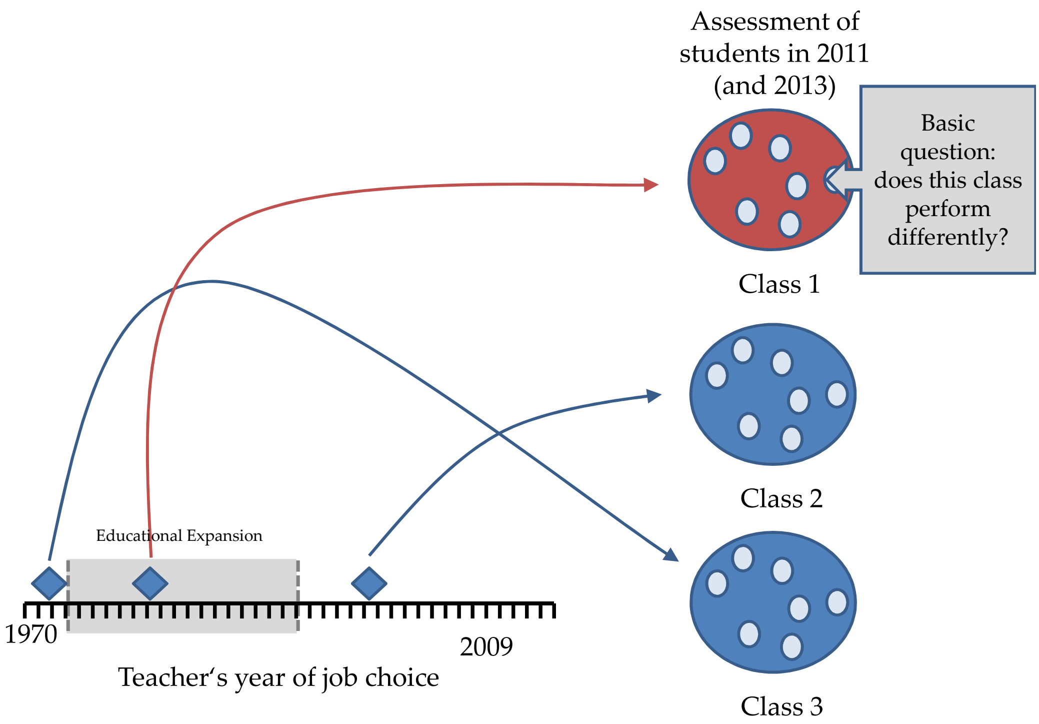


Figure 2: Illustration of the fixed effects setup

simply compare EETs with teachers who started their education after or before the educational expansion. The time scale at the bottom of Figure 2 illustrates this hypothetical clear temporal demarcation. However, the time of the educational expansion cannot be clearly defined. Yet, it can be exploited that the federal states in Germany have discretion over when, where, and to which extent to increase the capacity of the (secondary) educational system. Additionally, federal states decide on the curriculum in schools and in teacher training. Because of this institutional peculiarity, the mobility of teachers between federal states is low (Table A2 shows that nearly three quarters of teachers stay in the federal state where they graduated from high school). Consequently, I use the relative expansion of the teacher force at the federal state level to capture the part of the educational expansion that affected the job prospects of future teachers.

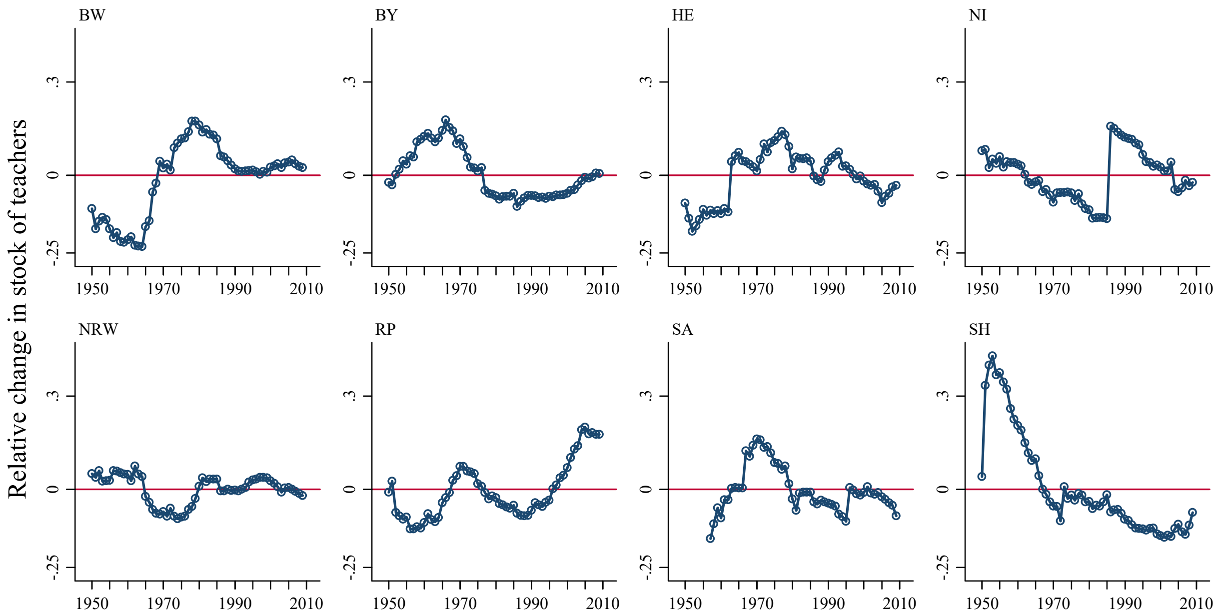


Figure 3: Relative changes in the stock of teachers by non-urban federal states over time
Notes: the time-series are residuals from a population-weighted regression of the stock of teachers on federal state and year fixed-effects. The data are based on administrative records and taken from Köhler and Lundgreen (2015).

This relative change in the stock of teachers over time and by federal state is depicted in Figure 3. In this graph, the differences in the timing as well as in the intensity with which the educational expansion was carried out are clearly visible. Each panel in Figure 3 depicts the West-German 'Flächenländer' (the urban federal states Berlin, Bremen, and Hamburg are excluded for the sake of clarity). The graph illustrates the different developments in the teacher market. If the teacher force of any given federal state grows faster relative to all federal states in a given year and faster than the own average growth rate, the relative changes plotted in Figure 3 are positive. Conversely, if the growth of the teacher force is lower than the trend in the federal states as well as the overall yearly change on the federal level, the relative change is negative. Another way to interpret the relative change in Figure 3 is by relating the number of (marginal) EETs to the number of teachers that were projected to be needed in the absence of the educational expansion, which is clarified in the next subsection.

Figure 2 also illustrates the general data structure that is exploited in the empirical approach. The three classes on the right-hand side of Figure 2 represent all fifth and ninth grades in the data. The pupils in those classes are subjected to objective tests on their math and reading performance. These test scores can be linked to teachers that teach the respective subject: German teachers are assigned to reading test scores and math teachers to math test scores. The educational expansion rate is merged to those teachers based on the federal state and the year (birth year plus 19) of their high school graduation. The effect of the educational expansion on students' learning outcomes may then be picked up by β_{FE} in the following regression:

$$y_{i_\tau l j_{st}} = \beta_{\mathrm{FE}} \ln(\#\text{teachers}_{st}) + \theta_s + \pi_t + \eta_l + \mu_\tau + \mathbf{X}'\boldsymbol{\rho} + \epsilon_{i_\tau l j_{st}} \tag{1}$$

where y measures the test scores of student i in year τ taught by teacher j at a school in state l who received his secondary school diploma in state s in year t. Because of the twofold fixed effects at the teacher level (θ_s and π_t), β_{FE} is essentially identified by relative deviations from the state-specific mean and the average yearly change across all federal states.[5] These deviations are exactly what is depicted in Figure 3. In addition to the teacher level fixed effects, student level fixed effects are also employed (η_l and μ_τ) to control for any persistent differences between the years and the federal states of the schools. Moreover, $\mathbf{X}$ may contain further covariates to possibly control for class composition, depending on the exact specification. In this fixed effects model, β_{FE} may pick up the effect of teacher quality on students' learning outcomes, if changes $\ln(\#\text{teachers}_{st})$ only capture the difference in teaching quality between EET and non-EET (see next subsection) with all else held fixed. However, one could still be concerned that skilled teachers have better opportunities to choose the school they teach in. Such a selection would confound β_{FE}.

To break the correlation between the initial skills of the students and teacher quality, variation between subjects (math and German) is exploited. Table 1 shows how this information helps to improve the identification. As every student has a German and a math teacher and is assessed in both reading and math skills, there are four possibilities for using the test score observations of a certain student (indicated by the gray-shaded cells). First, the math score is evaluated with respect to the exposure to the educational expansion (the relative changes depicted in Figure 3) of his math teacher. Second, reading scores and the exposure of the German teacher can be used. Both assessments are reflected in β_{FE}. This coefficient captures the direct effect of teacher quality plus, potentially, some school sorting effect. Moreover, also assessing across subjects can be informative: relating math scores to German teachers and reading scores to math teachers. Estimating Eq.

[5] Thus, it can also be termed a difference-in-differences model with continuous treatment. The reason why I refer to this model as 'fixed effects' is to clearly separate the wording from the difference-in-differences model that is employed later on.

Table 1: Setup of the difference-in-differences approach

	Math Scores	Reading Scores
Math Teacher	Treatment ($D = 1$)	Control ($D = 0$)
German Teacher	Control ($D = 0$)	Treatment ($D = 1$)

(1) using this cross-subject test score-teacher association yields the school sorting effect and potentially also the same spillover effect. In the absence of a spillover effect, the school sorting effect is identified and can be substracted out of β_{FE}. This can be directly done by defining a treatment and a control group (indicated by the treatment variable D) and by estimating the following model:

$$\begin{aligned} y_{i_{\tau lF} j_{tsF}} \quad = \quad & \alpha + \beta_{\mathrm{DiD}} \ln(\#\text{teachers}_{st}) \times D + \delta \ln(\#\text{teachers}_{st}) + \gamma D \\ & + \theta_{t \times D} + \theta_{s \times D} + \eta_l + \mu_\tau + \mathbf{X}'\boldsymbol{\rho} + \epsilon_{i_{\tau lF} j_{tsF}} \end{aligned} \tag{2}$$

Because this model differences out the school sorting effect, it is a difference-in-differences approach (DiD). The treatment group comprises students' test scores and teachers from the same subject and is indicated by the treatment indicator D taking the value 1. The control group, on the other hand, connects students' test scores and teachers between the subjects (math and German). This relation is indicated by $D = 0$. To facilitate interpretation, the fixed effects of the state and the year of the teacher's high school graduation are now interacted with D.[6] Finally, standard errors for β_{FE} and β_{DiD} are clustered on the federal state and year level of the teachers' high school exit exam since this is the level where the hiring of teachers occurs.

Besides a school sorting effect, this regression automatically purges all individual and also class and school fixed effects. If the assignment of German and math teachers to classes is mean-independent of teacher quality and of the relative, subject-specific skills of the class, the coefficient β_{DiD} identifies the causal effect (see Appendix A1.1 for a clear list of

[6] In the difference-in-differences equations as in (2) interpreting β_{DiD} as being identified from deviations from state and year-specific means would not work. To get these deviations, regress $\ln(\#\text{teachers}_{st}) \times D$ on the respective fixed effects (by the Frisch-Waugh-Lovell Theorem, a 'second stage' regression of y on ω_{st} and $\ln(\#\text{teachers}_{st})$ would yield the same coefficients as in Eq. (2) without interacted fixed effects):

$$\begin{aligned} \ln(\#\text{teachers}_{st}) \times D &= \mu_t \qquad + \eta_s \qquad + \omega_{st} \\ E\left[\ln(\#\text{teachers}_{st}) \times D\right] &= \delta_t \times \Pr(D) + \pi_s \times \Pr(D) + \epsilon_{st} \times \Pr(D) \end{aligned}$$

Applying the law of iterated expectations shows that the essential variation that identifies β_{DiD} is deflated by $\Pr(D)$. Using D-specific fixed effects adjusts for this deflation directly. Hence, interacted fixed effects are necessary in order to interpret β_{DiD} as deviations from the state-specific as well as the year-specific mean.

the identifying assumptions). Also, in the case of spillover effects, the school sorting effect is differenced out. Then β_{DiD} is a lower bound for the gross effect of teacher quality, since school sorting and spillover effects are both likely to be positive. However, the literature only finds weak evidence of the existence of spillover effects (Koedel, 2009). In robustness checks, however, I will scrutinize these spillover effects directly.

3.2 Theoretical mechanism

In response to the educational expansion, different individuals could have been encouraged to become teachers who also exhibit different career incentives. Why is that? As in every market, the labor market for teachers can also be characterized by two major forces, demand and supply. Regarding the former, the federal state s may project the demand for teachers in year t based on the expected number of academic track pupils, $\mathbb{E}_{st}P_{st}$. Also, the fraction of the teaching force that retires, δT_{st} may contribute to the demand for new teachers. In total, the overall demand for teachers can be expressed as $D_{st}(\mathbb{E}_{st}P_{st}, \delta T_{st})$. Because the federal states hire based only on how many students are enrolled or will enroll into the secondary educational system, supply-induced demand is unlikely to occur. Therefore, the demand can be seen as independent of the potential quality of teachers. It is exogenous to potential teachers.

Supply, on the other hand, is determined by the number of academic track graduates in year j and federal state s, as the job mobility between federal states is rather limited. Each individual within a cohort and a federal state has a net benefit of teaching $B(j_{st})$. This net benefit is the benefit of working as a teacher minus the benefit of working in the next best occupation. Hence, having the highest net benefit does not necessarily mean being the best teacher. It means that the skills or preferences of this individual are most teacher-specific. This benefit may depend on a vector of individual characteristics $\boldsymbol{S}_{j_{st}}$ of the potential teacher j_{st} that can be closely related to teacher quality $Q_{j_{st}}$. For instance, this vector may comprise intrinsic motivation to teach, specific teacher quality, and general skills among others. Thus, individuals with the highest benefit are most likely to be intrinsically motivated and have a high teacher quality. Similar to a Roy-type selection model of occupational choice (Roy, 1951), individuals will start teacher training based on this net benefit. But for individuals at the margin of becoming teachers, the decision may additionally depend on external market forces, such as the recruiting policy of the federal state. These individuals are less determined to join the teaching profession. Hence, extrinsic factors such as chances of eventually being hired as teachers, the prestige of the teaching job, or the relative salary are more important to those individuals.

Figure 4 plots the supply and demand forces. On the horizontal axis the share of academic track graduates in year t and federal state s with at most a certain teacher net benefit is depicted (for clarity, the scales are exaggerated). This share is mapped on the net benefit of being a teacher for all individuals in this cohort. Along the horizontal axis, the net benefit decreases. Thus, this supply function is equivalent to the quantile function of individuals having at most a certain net benefit. This is also called the inverted complementary distribution of the teacher net benefit: $q_{j_{st}} = (1 - F(B_{j_{st}}))^{-1}$.

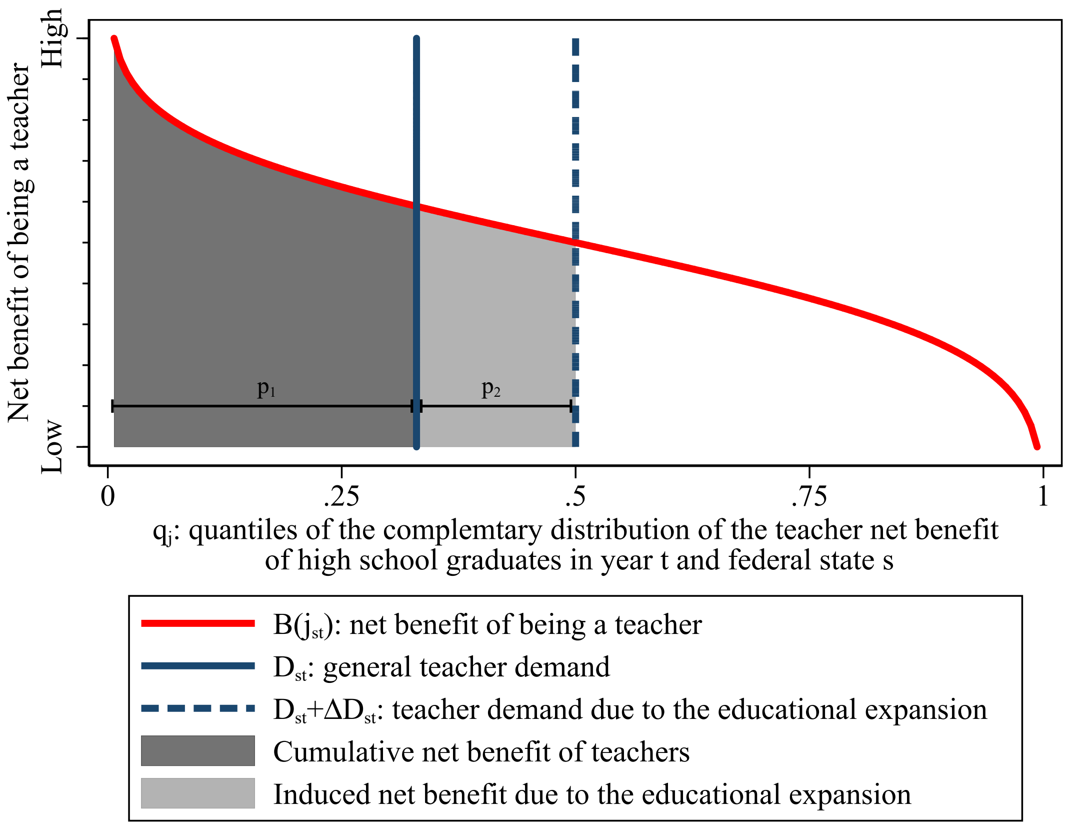


Figure 4: Possible impact of the educational expansion on the job market for teachers

In the absence of the educational expansion – which is targeted at increasing the share of each birth cohort with an academic track education – a fraction p_1 of each birth cohort can become teachers. This fraction depends on the demand for teachers D_{st}, which introduces external equilibrium factors to influence individual choices. Most likely, the individuals who become teachers are among those with the highest net benefit and implicitly exhibit those characteristics $\boldsymbol{S}_{j_{st}}$ that are better suited for being a good teacher. Note that D_{st} can also monotonously change from year to year in response to a constant fraction of teachers retiring or because the cohorts of students who transition to academic track education and those of high school graduates are constantly growing in the federal state.

In response to the educational expansion, there is an exogenous increase in D_{st}, denoted by ΔD_{st}. This has two notable consequences that outline the tradeoff between the quantity and quality of teachers. First, an additional fraction p_2 (the marginal teachers) of the high school graduate cohort that witnesses the demand increase for teachers in year t and federal state s decide to become teachers (the share of EETs amounts to $p_1 + p_2$ if $p_2 > 0$). The second consequence is that the average net benefit of all teachers – and therefore, most likely also the corresponding teacher quality – diminishes. In this model, the average net benefit of the p_1 teachers from a high school cohort in a federal state in normal years amounts to $\overline{B\left(D_{st}\right)} = \int_0^{p_1} B(j_{st})dF(q_{j_{st}})$ (depicted by the dark gray area in Figure 4) where $F(q_{j_{st}})$ is a uniform distribution (quantiles of a population are uniformly distributed). Accordingly, the average net benefit of those individuals who become teachers due to the educational expansion (marginal teachers) is: $\overline{B\left(\Delta D_{st}\right)} = \int_{p_1}^{p_1+p_2} B(j_{st})dF(q_{j_{st}})$ (indicated by the light gray area). The overall average net benefit (light and dark gray-shaded areas) of a teacher cohort t in federal state s that witnesses a teacher expansion (or contraction, $p_2 \neq 0$) can then be expressed as:

$$\underbrace{\overline{B\left(D_{st} + \Delta D_{st}\right)}}_{\text{Average net benefit of EETs}} = \overbrace{\overline{B(D_{st})}}^{\substack{\text{Average net benefit}\\ \text{of non-marginal}\\ \text{EETs}}} + \underbrace{\frac{p_2}{p_1 + p_2}}_{\substack{\text{Fraction of}\\ \text{marginal EETs}\\ \text{to all EETs}}} \overbrace{\left[\overline{B(\Delta D_{st})} - \overline{B(D_{st})}\right]}^{\substack{\text{Net benefit differential}\\ \text{between marginal and}\\ \text{non-marginal EEsT}}} \tag{3}$$

This expression explicitly shows how the average individual net benefit changes with respect to newly entering marginal teachers. The same effect applies not only to the net benefit but also to teacher quality if the benefit is monotonously related to the ability to teach (which is likely): $\overline{Q\left(D_{st} + \Delta D_{st}\right)}$. This equation is important in mainly two respects. First, ${p_2}/{(p_1+p_2)}$ is similar to the employed educational expansion rate as depicted in Figure 3. This rate is ${p_2}/{p_1}$. In the appendix, I show that the empirical results are insensitive to employing ${p_2}/{p_1}$, or ${p_2}/{(p_1+p_2)}$. Thus, it shows that the effect of the educational expansion on the labor market for teachers can be measured by the relative share of incoming teachers (rather than, for instance, the absolute number of teachers). As this is achieved by the log-specification, Eq. (3) justifies its use as the preferred specification in the empirical models of Eq. (1) and (2). Using $\ln(\#\text{teachers}_{st})$ mechanically adjusts the effect from all EETs ($p_1 + p_2$) to the marginal teachers (p_2, as a local average treatment effect adjusts the effect to the complying population) – the EET (light gray area) – and thus does not average the effect over all teachers in a particular cohort (light and dark gray areas). In this sense, one can think of this approach as also being an instrumental variables approach.

The second reason for why Eq. (3) is useful is for interpreting the results later. As outlined in the empirical strategy, I test whether p_2/p_1 is correlated with the test scores of students. If it is correlated, the effects in β_{FE} and β_{DiD} are given for the average change in teacher quality $[\overline{Q(\Delta D_{st})} - \overline{Q(D_{st})}]$ (if this is the exclusive driver behind the effect on student performance) averaged over all years and federal states (changes in $\overline{B(D_{st})}$ are captured by the fixed effects, π_t and θ_s in the regression models (1) and (2)). If this quality differential was observed, one could regress $[\overline{Q(\Delta D_{st})} - \overline{Q(D_{st})}]$ in a first state on $p_2/(p_1+p_2)$. Then, the reduced-form effect can be adjusted not only to the marginal teachers but also to a one-unit increase in teacher quality. These two features imply that the effects of the educational expansion can be precisely identified. In contrast, the effect of latent teacher quality on students' learning outcomes is a reduced-form effect (in terms of teacher quality) as teacher quality is unobserved.

4 Data

4.1 Sample selection and student-teacher linking

This study exploits the National Educational Panel Study (Blossfeld et al., 2011). The NEPS has a multi-cohort design and covers the educational trajectories of all individuals from six different stages of life. Specifically, I use the third (SC3) and the fourth (SC4) starting cohorts. SC3 comprises individuals that attended the fifth grade, whereas SC4 contains individuals from the ninth grade at the start of the school year 2010/2011. Compared to any survey data in Germany, the advantage of the NEPS is that it includes information on both the students and their teachers. The design of the questionnaire is equivalent across both cohorts. Hence, individuals and teachers from both starting cohorts can be pooled together in one sample.

The sampling population are all German fifth and ninth graders in 2010. In a first step, 234 schools are sampled (Skopek et al., 2012). All students in grades 5 and 9 from these schools are asked to participate in the survey. Since the NEPS is a panel survey, it follows these students as they move through the education system, including general education and occupational training. The survey also extends to the students' parents and the teachers in math, German as well as the class teachers. Teachers are interviewed once and can be linked to the respective class they teach. Information on teachers include year of birth, their high school graduation, their college education, retrospective determinants of their occupational choice and their attitude toward their job as a teacher.

Several restrictions need to be imposed on the data. From initially 1,206 teachers and 9,042 students that attend higher secondary schools in West Germany (the educational expansion did not take place east of the Iron Curtain, including East Germany), I restrict the sample to academic track schools. This group of students is high-skilled and mostly homogeneous in their abilities. Furthermore, I keep only teachers who either teach math and German (thereby dropping the class teachers). Both restrictions reduce the sample to 345 teachers and 4,259 students. Lastly, I restrict teachers to being younger than 60 years old as older teachers might already anticipating retirement . Therefore, I additionally drop 23 teachers. This means the oldest teachers in my final sample made the decision in 1970 to become a teacher, which is after the adjustment processes of teacher salaries and teacher training had finished.

Figure 5 shows the number of teachers in my sample by subject over time. There is approximately an equal amount of math and German teachers, and only a negligible minority teach both subjects. As is visible by the co-movement in the number of subject teachers over time, there is more variation over time than between subjects. In the NEPS teacher force, there are many teachers who graduated from high school (at age 19) in the 1970s. The 1980s are characterized by a saturated teacher force and relatively fewer hirings, which is also reflected in Figure 5. In the 1990s and 2000s (until 2005) the number of teachers in the sample increases again.

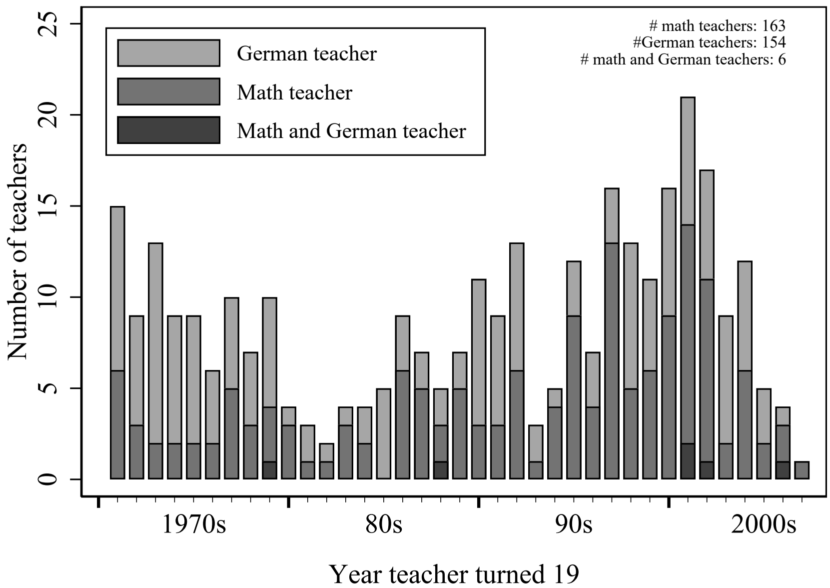


Figure 5: Number of teachers by year and subject
Notes: own calculations using NEPS data.

4.2 Test score data

THe unidimensional competence scores serve as the main outcome variables in reading and mathematics.[7] These scores have been assessed in tests conducted between November and January of a school year. As the school year usually starts in August, teachers can impact the test scores of their students through lessons in the first three to five months of the school year (on average, it is 3.72 months). Teachers cannot control the results of the test as these are conducted by trained NEPS interviewers. The scores are assessed by multiple choice questionnaires that every pupil has to fill in. The answers to these questions are aggregated by a weighted maximum likelihood estimation (WLE, Pohl and Carstensen, 2012). WLEs in the first wave are constrained to having a mean of zero. Values above zero therefore indicate abilities above average. This makes the scores comparable across the waves and cohorts. The variance of the WLE scores is not restricted.

Math competence score

Mathematical competence is targeted at measuring the "ability to flexibly use and apply mathematics in realistic situations" (Schnittjer and Duchhardt, 2015, p.2). Mathematical competence is assessed by 24 items in grade 5 and 22 items in grade nine on several domains.[8] For both grades, the test is designed to take 28 minutes in total. Examples of multiple choice questions include the following: "Mr. Brown owns a rectangular plot, which he wants to fence. After some calculations he buys 40m fence. The plot is 8m wide. How long is the plot?"

Reading scores

Understanding and using written texts is an important skill and a prerequisite for participating in cultural and social life (Gehrer et al., 2012). The reading score test is designed to measure those skills. As German lessons are designed to let students acquire the exact same skill, the reading score skills can be attributed to the domain of the German teacher. In order to accurately assess these skills, it is distinguished between five "text functions and associated text types" (informational texts, commenting or argumenting texts, literary texts, instructional texts, and advertising texts). Within the time of the test (also 28 minutes), the test participants are given the five types of texts ranging from informational to literary texts. Each type of text is associated with a different skill. Texts are adjusted to the lexical level, difficulty, and thematic orientation of the specific cohort and age level. The participants are ask to read a short text, before answering multiple choice questions Right after having read each text, .

[7] The data are cleaned from effects of position and order. This is achieved through a random assignment of the order of the two tests to respondents (Durchhardt and Gerdes, 2012).

[8] Quantity is captured by eight items, space and shape in total have five, change and relationships six and Data and chance five.

4.3 Descriptive statistics

Table 2 presents some descriptive statistics with respect to the educational expansion status of the teacher. For the sake of simplicity, the educational expansion rate p_2/p_1 is discretized at a threshold of zero. According to this definition, 2,203 students are taught by EETs, while 2,816 students have a non-EET.

Table 2: Descriptive statistics

	Educational expansion teachers: $p2/p1 > 0$		Non-expansion teachers: $p2/p1 \leq 0$	
	Mean	sd	Mean	sd
Test scores				
Reading	0.92	(1.13)	0.94	(1.14)
Math	1.01	(1.12)	1.08	(1.14)
Student characteristics				
Share female pupils	0.53	(0.50)	0.52	(0.50)
Teacher characteristics				
Share German teachers	0.49	(0.50)	0.55	(0.50)
Treatment, the relative expansion in the stock of teachers:				
Raw values:				
$\ln(\#\text{teachers}_{st})$ for German teachers	10.30	(0.61)	10.22	(0.49)
$\ln(\#\text{teachers}_{st})$ for math teachers	10.30	(0.80)	10.24	(0.60)
Effective variation: $p2/p1$ (plotted in Figure 3):[a]				
$p2/p1$ for German teachers	0.05	(0.04)	−0.03	(0.02)
$p2/p1$ for math teachers	0.03	(0.03)	−0.03	(0.02)
Class characteristics				
Class size	17.43	(5.94)	18.42	(5.35)
Minimum instructional time of teachers	3.58	(0.57)	3.86	(0.62)
General characteristics				
Share from SC4	0.64	(0.48)	0.53	(0.50)
Share from second wave among SC3 observations	0.60	(0.49)	0.53	(0.50)
Number of student-teacher-course-wave observations	2,203		2,816	

[a] Notes: This is the effective variation, which refers to the variation in $\ln(\#teachers_{st})$ when all other variables, most importantly federal state and year fixed effects, are held fixed: the residual of log stock on year and federal state fixed effects, which are relative changes in the federal state-specific stock of teachers from the general expansion trend across all federal states.

Educational expansion teachers teach students with a worse test score (0.92 vs. 0.94 for reading and 1.02 vs. 1.08 for math) – a first descriptive indication of an effect. The gender of the students is balanced between EET and non-EET. German teachers are less likely to be classified as EET according to my definition. Potentially, this is because math teachers possess skills that make them react more sensitively to the changed career incentives of the educational expansion. The next four characteristics refer to the educational expansion rate. It is shown as its raw values (the log stock of teachers) and as the effective variation (demeaned by year and federal state fixed effects). These measures are presented separately for math and German teachers.

The average class size differs slightly between EET and non-EET (17.4 vs. 18.4). The instructional time (time from start of the school year to the assessment of the test score) also varies slightly according to the educational expansion status of the teacher. In the overall sample, slightly more students are in the initial ninth grade (SC4). The students in this grade have a higher chance of being taught by an EET. Within the initial fifth grade (SC3), 56 percent of the observations come from the second wave (all observations from SC4 are assessed in the first wave). This statistic also varies somewhat by the educational expansion status of the teacher. Although the sample appears to be slightly imbalanced in these respects, the empirical strategy and the robustness checks rule out that imbalances between cohorts and waves can carry over to the identification of the main effects.

5 Results

5.1 Effects on students' learning outcomes

Table 3 presents the estimation results from Eq. (1), the baseline fixed effects results specification by subject. It is a first step in clarifying whether individuals were encouraged to become teachers by the educational expansion and are now teaching students that today perform differently at school. The first line of Table 3 shows the association between the change in the stock of teachers in the year and the federal state of high school graduation and the respective test score of the pupils that they taught in the survey year. The first three columns refer to math teachers and the associated math score of their pupils, the last three columns are results for German teachers and the reading score of their pupils. On average, the math competence score is 0.0127 points lower for every 1 percent that the stock of teachers increased relative to the overall trend in the year the teacher turned 19 and decided on his future job (as reflected by p_2/p_1). Two things are worth noting: first, the result is non-negligible in magnitude and suggests that teachers play an important role. Why is the coefficient plausible? The mean effective variation that identifies β_{FE} (the mean absolute deviation of the residual of a regression of $\ln(\#\text{teachers}_{st})$ on all the controls) shows that the mean change in the stock of teachers was 4.33 percent on average. Multiplying β_{FE} with this variation and dividing by the standard deviation in

Table 3: Fixed effects results for mach and reading competence

	Math teacher – math competence			German teacher –reading competence		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\#\text{teacher}_{st})$	-1.237^{**}	-1.295^{**}	-1.382^{***}	-0.382	-0.650	-0.743
	(0.536)	(0.510)	(0.509)	(0.500)	(0.456)	(0.464)
Further condition on:						
_ Cross-subject competence score		✓	✓		✓	✓
_ Federal state of school FE			✓			✓
Observations	2,713	2,620	2,620	2,437	2,399	2,399
Number of teachers	168	168	168	158	158	158

Federal-state-by-year-level clustered standard errors in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. All columns refer to a separate regression with additional federal state and year fixed effects plus all effects indicated.

math competence indicates that 5.3 percent[9] of a math score standard deviation can on average be attributed to the educational expansion (if interpreted as causal). As I will try to demonstrate below, this magnitude fits well into what previous studies found. The second notable point is that the effect is robust toward the inclusion of important control variables that may mitigate the role of school selection: including reading competence is supposed to capture the general ability of the student whereas state of school fixed effects should control for persistent migration patterns of teachers within Germany. Because the results are robust toward the inclusion of these fixed effects, migration of teachers (shown in Table A2) does not affects the results.

For reading competence, the results are somewhat different, although the direction of the effect is unchanged. Having a teacher that was gradually exposed to a higher degree of the educational expansion – as measured by a 1 percent increase in the relative change in the stock of teachers – goes along with having a 0.0038–0.0074 lower score in reading competence depending on the specification. Applying the same calculation as above yields the fraction of a standard deviation in reading scores that can be attributed to the educational expansion (again, a causal interpretation) shows that this fraction amounts to 2.79 percent. Note, however, that none of these results are significant at the 5 percent level. Moreover, recall that the finding of smaller effects on reading competence is in line with the literature where, for instance, Nagler et al. (2015) also find smaller effects of recession teachers on the reading value-added measure of their students. Also, Chetty et al. (2014a) report a smaller value-added transmission on reading compared to math scores. In the context of this paper, this finding can be due to two reasons. First, German teachers may

[9] Calculation: $1.3827[\text{coefficient}] \times 4.33[[\text{mean absolute deviation in }\%]/1.13[\text{sd of test score}]$.

generally have a lower leverage on reading scores whereas the math score might better capture what is taught in the lessons. Second, the German teachers might have reacted differently to the educational expansion such that the effect on teacher quality is not that pronounced. One reason for this can be the potentially better outside option for math teachers.

How likely is it that these effects are attributable to the teacher and not to some unobserved class, school, or individual characteristics? To answering this important question, I now turn to the difference-in-differences estimation outlined in equation (2). Its results are presented in Table 4. These are the main results of the paper, since it comes closest to answer the question – what is the effect of teacher selection induced by the educational expansion on the learning outcomes of today's pupils. To approach an answer, I first pool data from all the cells of Table 1 into one comprehensive sample. As a result, I have one pupil by test score observation by teacher (see an example data set in Table A3), but every pupil can now appear in the sample up to four times. This approach allows me to use information on all teachers and students simultaneously. To adjust standard errors to this restructuring, standard errors remain clustered on federal state by year level as before and throughout the whole analysis. In Table 4 the main coefficients are presented, with subsequently added control variables as one moves from the left to the right columns. The

Table 4: Main results – impact of the educational expansion on students' test scores

	Competence scores							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln(\#teachers_{stF}) \times D$	-0.822^{**}	-0.822^{**}	-0.816^{**}	-0.966^{**}	-0.966^{**}	-0.968^{**}	-0.969^{**}	-0.966^{**}
	(0.382)	(0.382)	(0.0383)	(0.378)	(0.378)	(0.381)	(0.381)	(0.381)
$\ln(\#\text{teachers}_{stF})$	0.332	0.341	0.321	0.229	0.289	0.033	0.058	0.335
	(0.413)	(0.417)	(0.410)	(0.358)	(0.347)	(0.310)	(0.307)	(0.399)
Subject FE		✓	✓	✓	✓	✓	✓	✓
Gender			✓	✓	✓	✓	✓	✓
School state FE				✓	✓	✓	✓	✓
Cohort FE					✓	✓	✓	✓
Wave & class FE						✓	✓	✓
Test month FE							✓	✓
State specific trends								✓
Observations	10,330							
Number of pupils	6,772							
Number of teachers	322							

Federal-state-by-year-level clustered standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$. Baseline regression equation is shown in (2). All columns refer to a separate regression with additional Federal State and year fixed effects plus all effects indicated.

main and most important effect listed in the first line ($\ln(\#\text{teachers}_{st}) \times D$). It captures the additional effect of the educational expansion of teachers that teach the corresponding subject measured by the outcome variable (math competence for math teachers and reading competence for reading teachers). This effect is significant and robust toward the inclusion of further fixed effects (columns 2–8): explicit subject fixed effects do not change the result (column 2; as they are implicitly incorporated in Eq. (2)), neither do the characteristics of the teachers (column 3). Including state of school fixed effect slightly inflate the effect (column 4), whereas cohort, wave, class, test month fixed effects nor even state-specific trends impact the coefficient any further. Causally interpreting this effect means: every 1 percent of a higher relative demand for teachers would attract teachers that – on average – reduce the subject-specific test scores of their pupils by 0.00822 to 0.00966. Conducting the same exercise as above and taking the mean effective variation that identifies the effect for $\ln(\#\text{teachers}_{st}) \times D$ – which in this setting amounts to 2.38 percent – shows that 2.02 percent of the overall standard deviation can on average be attributed to the educational expansion.[10] The difference between this fraction and the average effects of the FE model in math and reading (5.3 for math and 2.8 for reading – roughly equal to 4 percent) can hence be attributed to a selection effect that the first analysis was not able to control for.

How do the effects place themselves in the literature? Chetty et al. (2014a) use an event study of teachers who move between schools as a natural experiment to assess the impact on the test scores of the newly taught students. They find that test scores are raised by 3.5 percent of a sd because of the entry of a teacher from the top 5 percent of the teacher value-added distribution (as assessed by data on previous years). On the one hand, Eq. (3) shows that the effects in β_{FE} and β_{DiD} are already adjusted to the educational expansion teachers (by p_2/p_1, see Eq. (3)). On the other hand, it is not adjusted to the average quality differential between marginal and non-marginal EETs. Because this differential is most likely to be lower than between a teacher from the top 5 percent versus an average teacher, the β_{FE} and β_{DiD} needs to be inflated. This fact puts my results even more into the range of the findings of Chetty et al. (2014a). The results presented here are in that sense reduced-form effects, since I am not able to normalize them using value-added measures (as the second stage of a two-stage-least-squares estimation would do).

If I expect the same effects as in Chetty et al. (2014a) to operate in my data (0.14sd for math and 0.10 for English:[11] 0.12 on average), I can back out a first stage: the effect of an expanding teacher force on teacher quality (the quality differential in Eq. (3)). In this case, every 1 percent increase in p_2/p_1 would induce individuals to become teachers such

[10] Calculation: $-0.966[\text{coefficient}] \times 2.38[\text{mean absolute deviation in }\%]/1.14[\text{sd of test score}]$.

[11] The scores are normalized on a one-sd increase in the teacher value-added.

that the value-added of the whole teachers' cohort is increased by 0.0805sd.[12] Also, the literature offers estimates on such a "first stage." Nagler et al. (2015) aim at estimating the effect of recessions on teacher quality, which may be roughly comparable to this setting. They find that due to a recession, the teacher value-added increases by 0.11sd in math and 0.05sd in reading for recession teachers. On average, this is equivalent to the back-of-the-envelope calculation that also yields 0.08.

5.2 Assessing the validity of the estimates

Threats to the identifying assumptions

To check that the overall effects are not driven by anything but the causal effect of the subject teacher on the subject test score, I present two complementary pieces of evidence in Table 5.

Table 5: Robustness checks – placebo regression and predicting parental characteristics

	Placebo regression		Parental characteristics		
	Math teacher – reading score	German teacher – math score	log HH income	Edu. years: mother	Edu. years: father
	(1)	(2)	(3)	(4)	(5)
$\ln(\#\text{teachers}_{st})$	0.079	0.268	-0.085	-0.236	-0.110
	(0.514)	(0.516)	(0.399)	(0.353)	(0.360)
Observations	2,713	2,437	2,361	4,079	2,749
Number of teachers	168	158	226	343	315

Federal-state-by-year-level clustered standard errors in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.
All columns refer to a separate regression with federal state and year fixed effects.

First, I present a placebo regression where I assign to each teacher the cross-subject test score; hence, reading scores to math teachers and vice versa (put differently, regression model (1) is estimated within each of the light gray cells in Table 1). Results of this placebo regression are presented in the first two columns of Table 5. If at all, having a math teacher

[12]The exact calculation looks like this:

$$\underbrace{\text{Second Stage}}_{\text{from Chetty et al. (2014a)}} = \frac{\overbrace{\text{Reduced Form}}^{\text{from Table 4}}}{\text{First Stage}}$$

$$\Leftrightarrow \text{First Stage} = \frac{0.00966\left[\frac{\text{Test score}}{1\text{ \% increase in \# teacher}_{st}}\right]}{0.12\left[\frac{\text{Test score}}{\text{Teacher value-added}}\right]} = 0.0805\left[\frac{\text{Teacher value-added}}{1\text{ \% increase in \# teacher}_{st}}\right]$$

who took up the profession because of the educational expansion raises the reading competence scores of his students (column 1). Similarly, this kind of teacher in German does not decrease his student's math competence score (column 2). This finding is consistent with the notion that teachers affect the test score mainly in the subject they teach. Thus, there is not much evidence of either a school selection effect or a spillover effect.

Second, an implicit assumption of the regression models (1) and (2) is that – conditional on all controls, foremost the fixed effects – everything apart from the educational expansion rate of the teacher is held fixed, even potential factors that are not incorporated in the regression (see Pei et al., 2017 for details). To test for this, I consider potentially important predictors for students' learning outcomes: their socio-economic background measured by the log household income of the parents as well as the years of education of both the fathers and mothers. If, in a pooled regression (math and German teachers), the teachers' educational expansion rate at the time of his high school graduation is able to predict the parental background of the teachers' students, at least part of the effect could be put into question. In this case, it would not be sufficient to control for the parental background, as further important variables that are still left out of the regression are easily conceivable. Results of this analysis are presented in the last three columns of Table 5. It shows that changes in $\ln(\#\text{teacher}_{st})$ have neither the power to predict the household income of the student (column 1), nor years of education of the mothers (column 2) or fathers (column 3). Hence, both supplementary analyses support a causal interpretation of the effects of β_{FE} presented in Table 3. It should be noted, however, that math teachers have a marginal impact on reading competence – even more so vice versa. Additionally, EET also teach pupils from a marginally more adverse background.

A caveat may be teacher non-response if it is correlated with the educational expansion rate. Table A7 shows that teachers who are willing to provide some background information also teach students that score higher in the math and reading tests. However, this effect disappears once it is conditioned on school fixed effects. This finding suggests that school principals and peer pressure may mainly enforce participation. Using the main specification (2), the consent of the subject teachers is not at all able to predict the scores in his subject. Thus, teacher non-response is an argument to prefer the difference-in-differences over the fixed effects model.

A further concern – that may apply to the fixed effects as well as to the difference-in-differences setup – might be the sensitivity of the effects with regard to the assignment year. Figure A3 evaluates the sensitivity of the effect with regard to changes in the assignment year. As it reveals, the conclusion and interpretation of the results does not depend on the exact assignment year. The effects are stable over the range where individuals usually make their job decision. Outside of this range (for instance, before age 15 and after age 25) effects disappear. Lastly, the results are insensitive to the size of the class that the teacher teaches (Table S.A1) and the class size and fraction of students with valid test scores are uncorrelated with the educational expansion rate of the teacher (Table S.A2).

The expansion in tertiary education and its relation to the quality of teacher training As Kamhöfer et al. (2017) demonstrate, the educational expansion also massively affected the university landscape of Germany (from 1962 to 1990, the number of universities doubled). Hence, it is legitimate to ask whether the potential teacher quality differential underlying the main results stems from a difference in the quality of the teacher training in newly opened universities. Table 6 therefore presents evidence on whether quality differentials at the university level are a relevant driving force. To check whether factors on the university side are driving the results, I rerun the most saturated specification from Table 4 (presented again in column 1 of Table 6) and further add university fixed effects (column 2).

Table 6: Driving force behind effect

	β_{DiD}		
	(1) [a]	(2)	(3)
$\ln(\#\text{teachers}_{stF}) \times D$	-0.966^{***}	-0.727^{*}	-1.231^{***}
	(0.381)	(0.368)	(0.381)
$\ln(\#\text{teachers}_{stF})$	0.335	0.319	0.407
	(0.399)	(0.435)	(0.408)
Teachers' university fixed effects		✓	
Teachers from new universities dropped			✓
Observations	10,330		9,156
Number of teachers	322		281

Federal-state-by-year-level clustered standard errors in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.
[a] As shown in column (8) of Table 4

Although the magnitude of the effect shrinks by about one quarter in absolute terms, the effect remains significant and economically relevant even after absorbing a potentially high fraction of the identifying variation. Thus, the result indicates that heterogeneity in university quality only explains a small fraction of the effect. But openings can also lead to a selection of high-ability individuals becoming teachers. To check this, I drop teachers that graduate from new universities and re-estimate Eq. (2). The resulting estimate is higher and thereby provides some evidence that university openings generally induced teachers of a higher quality to enroll in teacher training.

5.3 Detecting teacher selection in the characteristics of teachers

So far, I looked at whether teachers have a different ability (i.e., teacher quality) to raise the test scores of their students with respect to different degrees of their exposure to the educational expansion. Although this is considered to be the ultimate measure of teacher quality (see, e.g., Hanushek and Rivkin, 2006 or Chetty et al., 2014b), one can still ask whether the teachers not only have a better quality but also different characteristics that are correlated with quality (Jackson et al., 2014). This serves two purposes. First, if I found effects, this would strengthen the credibility of the main effects on test scores. And second, it is important for tailoring future policies, since hiring decisions or enrollment conditions for prospective teachers may be based on characteristics that correlate with teacher quality. The NEPS data set provides additional information on teachers. In addition to the birth year and the federal state of high school graduation that was used throughout the analysis, the data also includes the grade of high school and university graduation. In addition, the data contains subjective indicators that are targeted to retrospectively portray aspects of the reasons why they became teachers. Ten questions in the questionnaire for teachers try to capture these aspects. Teachers have to assess the relative, subjective importance of each aspect on a four-point Likert scale (ranging from very unimportant, 1, to very important, 5). For two reasons, it may be suboptimal to present estimates on all 10 domains. First, multiple testing may be a concerning issue, since one cannot determine at which domain to expect an effect and on which not a priori. Second, teachers may differ generally in their answer patterns. For instance, low-quality-teachers may place a higher importance on all domains in general. High-quality teachers may tend to place less weight on all domains but relatively more on those that correlate with intrinsic motivation. Those two opposing patterns may then confound the overall effect.

I therefore conduct a factor analysis that serves to detect these patterns. This is similar to Rockoff et al. (2011) who employ variables on cognitive skills. Because I expect two latent factors to be inherent in the answer patterns – namely intrinsic and extrinsic motivation – I opt for a principal component analysis with two factors.[13] For the 10 questions, the resulting two factor loadings are plotted in Figure 6. The horizontal axis maps the first dimension and the vertical axis the second factor loading. The loadings on the first domain are all positive. This can be ascribed to a general positive correlation between all of these subjective questions. This general correlation is purged out of the second loading. Therefore, it may be more informative for the analysis. Indeed, the second domain clearly shows that the variables form two clusters. Specifically, the importance of leisure, salary,

[13]Principal component analysis simply transforms p-dimensional data into $m < p$ dimensional data, where p is the number of principal components along which the data varies most. Technically, the first component is a summary score of the data $PC_1 = \phi_{11}x_1 + \phi_{21}x_2 + \cdots + \phi_{101}x_{10}$ and ϕ_{i1} are the factor loadings of the first component. The ϕ's are chosen such that they maximize the sample variance of PC_1 under the constraint that $\sum_{i=1}^{10} \phi_1 i^2 = 1$. The second principal component PC_2 again maximize the variance of the data, but with the additional condition that PC_2 is orthogonal to PC_1.

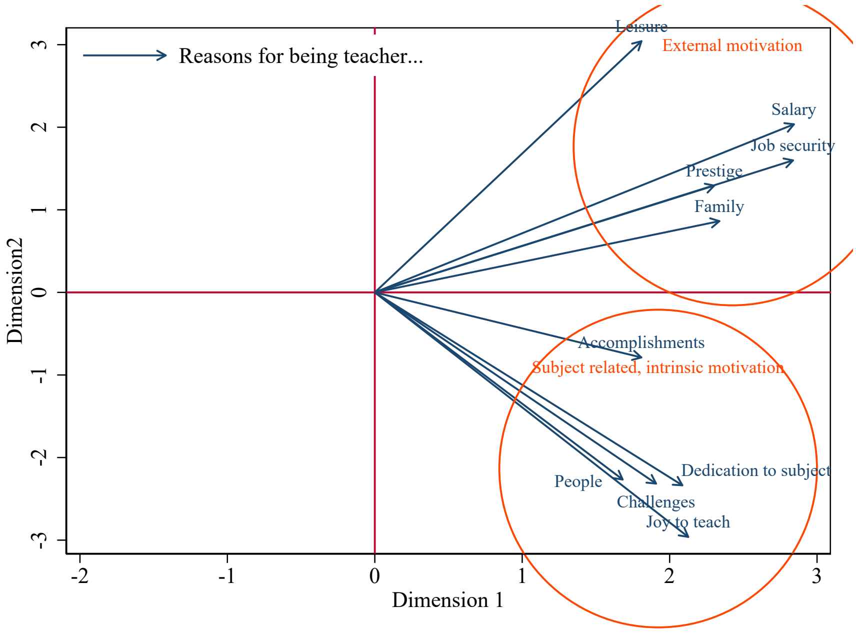


Figure 6: Cluster analysis of aspects teachers' job choice

Notes: the graph (biplot) plots the factor loadings resulting from a principal component analysis with two components on 10 variables that capture the aspects of the job choice of the teachers.

job security, the prestige of the job, and being able to reconcile the job with a family life form one cluster (positive factor loading). Since all those domains are not specific to the teacher profession, I refer to these variables as those reflecting external motivation. The remaining variables have a negative factor loading. These variables comprise the joy to teach, the challenges of the job, being around people, the dedication to the subject and to accomplish certain goals in the job. The common feature of these variables is that they are all job-related. Hence, this cluster reflects intrinsic motivation. These two clusters are present in the latent correlation of the variables. It is crucial whether the scores formed by those factor loadings are affected by the exposure to the educational expansion ($p2/p1$). If the scores and $p2/p1$ were correlated, this would indicate that EETs have a different kind of motivation.

The bar plot in Figure 7 presents evidence on this. Each bar represents the effect of $\ln(\#\text{teacher}_{st})$ on a respective outcome variable (indicated by the label below each bar). The sample is equivalent to the DiD regression. As before, standard errors remain adjusted to the federal state-year level. The first bar shows that the higher p_2/p_1, the worse the grade of academic track high school graduation (Abitur). Hence, this effect indicates that teachers with worse high school grades take up the teaching profession at times of high demand for teachers. The effect, however, fails to be significant at the 10 percent level. Does this hint at the lack of statistical precision or point to negligible economic meaning? Table A6 tries to shed light on this question by comparing the coefficients of a teacher-level regression of different samples. It turns out that the coefficients are stable,

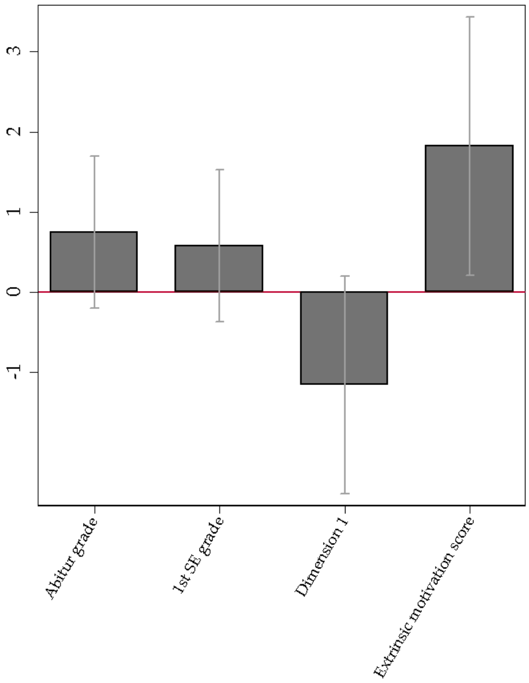


Figure 7: The educational expansion and teachers' characteristics

Note each bar depicts the effect of ln(#teachers$_{st}$) on the outcome indicated by the label below the bar. The sample is equivalent to the difference-in-differences regression with 322 teachers, standard errors are adjusted on the teacher level. Confidence bands indicate the 90% confidence level.

irrespective of whether academic track teachers from other subjects without assignable student test scores are included (column 2) or middle school (Realschule) teachers are further added (first column 1). But statistical precision increases by adding more teachers. These results for the German educational expansion are in contrast to findings for the US where no powerful predictors of teacher quality are identified (Jackson et al., 2014).

Returning to Figure 7, this figure further shows that the effect on the high school grade also propagates to university. Here, EETs have marginally worse grades. Beyond grades, is there evidence that teachers affected by the education expansion have a different work ethic? The third and fourth bars shed light on this by analyzing the principal component summary measures. The former shows that EETs tend to generally place significantly less importance on all domains captured by the questions because the first domain places almost equal and positive weight on all the domains. One explanation for this effect is a potentially different reference point of those teachers. Yet, distinguishing these questions as suggested by the second dimension is more informative. On this dimension, I find that there is a positive effect for EETs. This means that EETs place significantly more weight on questions with a positive weight (the external motivation to be a teacher) and less on those with a negative weight (the intrinsic cluster of the questions). This finding suggests that EETs have a slightly shifted work ethic from intrinsic to extrinsic motivation, which is compatible with the the main-effect: EETs may not put as much effort into raising the test scores of the students because they do not gain their motivation from it.

6 Conclusion

This paper shows that more teachers do not mechanically lead to smarter students. It thereby emphasizes an important mechanism of quickly expanding public institutions: focusing on quantitative aspects may deter quality, all else equal. This message can also be important for today's objective to increase the scale of public institutions, such as the current expansion of daycare facilities for children (BMFSFJ, 2015).

Using one of the major social changes in the past 60 years, the educational expansion, I test whether this social change attracted individuals with a different quality to eventually become a teacher. With the help of a simple expression of how the group average of teacher quality changes in response to newly entering teachers, the effect can be placed into the literature of broadly related studies. In a baseline fixed effects setup the impact of the educational expansion on teachers is separated from teachers' experience and federal state-specific effects. To take care that no school selection effects impact the results, I estimate a between-subjects difference-in-differences model.

The evidence I get from this approach suggests that the average effect of the educational expansion, which caused teacher quality to diminish, was roughly 2 percent of a standard deviation in students' test scores (math and reading). Comparing this ("reduced-form") effect to existing studies on teacher selection (e.g., Nagler et al., 2015 who provide a surrogate of a "first stage") and with the effect of teacher quality on students' test scores (Chetty et al., 2014a, a "second stage") the results of this paper are well-placed into the existing literature on the US. Thus, the scope of the effects on the students in this study are likely to also extend to labor market performance in adulthood. These results are further substantiated by the findings that teachers who are selected because of the educational expansion performed better at high school (though not at university) and have a slightly different work ethic that is based more on extrinsic rather than intrinsic motivation.

Potential policy implications are non-trivial, since not expanding the higher secondary education would not have been a solution either. Nonetheless, policy could very well have reduced its demand for teachers while sticking to the provision of a sufficient amount of spots for students. Taking the evidence of this paper together with a further characteristic of the educational expansion – the student-teacher ratio that declined at the same time in Germany (depicted in Figure A1a) – the educational policy departments of the federal states may have attenuated this tradeoff between teachers and students' learning outcomes by not simultaneously pushing down the student-teacher ratio. Either granting more pupils access to higher secondary education while increasing the student-teacher ratio at the same time or focusing more on investing in quality. For instance, this can be done by improved teacher training or via a more selective process of hiring teachers.

References

Ashraf, N., Bandiera, O., and Lee, S. (2014). Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery. STICERD - Economic Organisation and Public Policy Discussion Papers Series 54, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Bacolod, M. P. (2007). Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality. *The Review of Economics and Statistics*, *89*(4), 737–751.

Blossfeld, H.-P., Roßbach, H.-G., von Maurice, J., Schneider, T., Kiesl, S. K., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S., Prenzel, M. S., et al. (2011). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Age*, *74*(73), 72.

BMFSFJ, Bundesministerium für Familie, Senioren, Frauen und Jugend. (2015). Fünfter Breicht zur Evaluation des Kinderförderungsgesetzes – Bericht der Bundesregierung 2015 über den Stand des Ausbaus der Kindertagesbetreuung für Kinder unter drei Jahren für das Berichtsjahr 2014 und Bilanzierung des Ausbaus durch das Kinderförderungsgesetz. Tech. rep.

Bölling, R. (1983). *Sozialgeschichte der deutschen Lehrer: Ein Überblick von 1800 bis zur Gegenwart*. Vandenhoeck & Ruprecht.

Britton, J., and Propper, C. (2016). Teacher Pay and School Productivity: Exploiting Wage Regulation. *Journal of Public Economics*, *133*(Supplement C), 75 – 89.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–2679.

Dahrendorf, R. (1965). *Bildung ist Bürgerrecht: Plädoyer für eine aktive Bildungspolitik*. Nannen Verlag.

Durchhardt, C., and Gerdes, A. (2012). NEPS Technical Report for Mathematics – Scaling Results of Starting Cohort 3 in Fifth Grade. Tech. rep., NEPS Working Papers No.19.

Figlio, D. N. (1997). Teacher Salaries and Teacher Quality. *Economics Letters*, *55*(2), 267–271.

Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2012). The Assessment of Reading Competence (Including Sample Items For Grade 5 and 9). Tech. rep., NEPS research data – Leibnitz Institute for Educational Trajectories.

Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation using Micro Data. *American Economic Review*, *61*(2), 280–288.

Hanushek, E. A., and Rivkin, S. G. (2006). Teacher Quality. *Handbook of the Economics of Education*, *2*, 1051–1078.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavit, A. (2010). The Rate of Return to the HighScope Perry Preschool Program. *Journal of Public Economics*, *94*(1-2), 114 – 128.

Jackson, C. K., Rockoff, J. E., and Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics*, *6*(1), 801–825.

Jürges, H., Reinhold, S., and Salm, M. (2011). Does Schooling Affect Health Behavior? Evidence from the Educational Expansion in Western Germany. *Economics of Education Review*, *30*(5), 862–872.

Kamhöfer, D., Schmitz, H., and Westphal, M. (2017). Heterogeneity in Marginal Non-monetary Returns to Higher Education. *Journal of the European Economic Association*, *forthcoming*.

Koedel, C. (2009). An Empirical Analysis of Teacher Spillover Effects in Secondary School. *Economics of Education Review*, *28*(6), 682 – 692.

Köhler, H., and Lundgreen, P. (2015). General Secondary Schools in the Federal Republic of Germany from 1949 to 2010.

Lakdawalla, D. (2001). The Declining Quality of Teachers. NBER Working Papers 8263, National Bureau of Economic Research, Inc.

Loeb, S., and Page, M. E. (2000). Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-pecuniary Variation. *The Review of Economics and Statistics*, *82*(3), 393–408.

Lundgreen, P., and Schallmann, J. (2013). *Datenhandbuch zur deutschen Bildungsgeschichte*, vol. XI: Die Lehrer an den Schulen in der Bundesrepublik Deutschland 1949-2009. Vandenhoeck & Ruprecht.

Nagler, M., Piopiunik, M., and West, M. R. (2015). Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness. NBER Working Papers 21393, National Bureau of Economic Research, Inc.

Osterroth, F., and Schuster, D. (2000). Chronik der deutschen sozialdemokratie.

Pei, Z., Pischke, J.-S., and Schwandt, H. (2017). Poorly Measured Confounders are More Useful on the Left Than on the Right. Tech. rep., National Bureau of Economic Research.

Picht, G. (1964). Zwei Millionen Schüler mehr – Woher sollen die Lehrer kommen? *Christ und Welt*, *17*(3), 3.

Picht, G. (1965). *Die deutsche Bildungskatastrophe*. Deutscher Taschenbuch Verlag München.

Pohl, S., and Carstensen, C. H. (2012). NEPS Technical Report for Mathematics – Scaling the Data of the Competence Tests. Tech. rep., NEPS Working Papers No.19.

Rockoff, J. E., Jacob, B. A., Kane, T. J., and Staiger, D. O. (2011). Can You Recognize an Effective Teacher When You Recruit One? *Education Finance and Policy*, *6*(1), 43–74.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, *3*(2), 135–146.

Schnittjer, I., and Duchhardt, C. (2015). Mathematical Competence: Framework and Exemplary Test Items. Tech. rep., NEPS research data – Leibnitz Institute for Educational Trajectories.

Siegler, B. (2012). The Effect of University Openings on Local Human Capital Formation: Difference-in-Differences Evidence from Germany. Tech. rep., BGPE Discussion Paper.

Skopek, J., Pink, S., and Bela, D. (2012). Starting Cohort 3: 5th Grade (SC3) – SUF Version 1.0.0 Data Manual. Tech. rep., NEPS research data – Leibnitz Institute for Educational Trajectories.

Zimmerman, S. D. (2014). The Returns to College Admission for Academically Marginal Students. *Journal of Labor Economics*, *32*(4), 711–754.

A1 Appendix

A1.1 Assumptions of the difference-in-differences model

The underlying assumptions of this approach are threefold. The first assumption is that teacher quality matters similarly for math courses as it does for German courses: $Y^{\text{Math}}\left(Q^{\text{Math}}\right) \approx Y^{\text{Reading}}\left(Q^{\text{German}}\right)$, where $Y^u(Q^v)$ refers to the potential test score of a student in subject u which might depend on the latent teacher quality Q of a teacher who teaches the student in subject v. This assumption is important for the interpretation of the effect.[14] In the same vein, the second assumption rules out which effect I do not expect to see. If $u = v$ (a teacher in a certain subject can only affect the test scores of her students in the same subject) I expect to see an effect else it can be ruled out.

$$Y^{\text{Reading}}\left(\mathrm{Q}^{\text{Math}}\right) = Y^{\text{Reading}} \quad \wedge \quad Y^{\text{Math}}\left(\mathrm{Q}^{\text{German}}\right) = Y^{\text{Math}}$$

Those two assumptions allow me to precisely define a treatment indicator that indicates whether the teacher's subject ($v_j \in \{1,2\}$) is the same as the test score under consideration ($u_i \in \{1,2\}$).

$$\begin{aligned} D &= \begin{cases} 1 \quad \text{if} \quad (\overbrace{\text{test=Math}}^{u_i=1}) \quad \wedge \quad (\overbrace{\text{teacher=Math}}^{v_j=1}) \\ \qquad \vee\ (\underbrace{\text{test=Reading}}_{u_i=2}) \quad \wedge \quad (\underbrace{\text{teacher=German}}_{v_j=2}) \\ 0 \quad \text{else.} \end{cases} \\ &= \mathbb{1}\left(u_i = v_j\right) \end{aligned}$$

Also, these assumptions enable us to redefine the potential outcomes as Y^1, Y^0 in order to reconcile it with the treatment indicator.

The third assumption is actually most crucial for identification, since it states which variation in the response variable can be causally attributed to variation in the gradual changes in the measure of the educational expansion. To be more precise, I assume that the quality differential of any pair of math and German teachers in the same class is independent of the potential test scores of their pupils:

$$\left(Y^1(\mathrm{Q}^1) - Y^1(\mathrm{Q}^0)\right) \perp\!\!\!\perp \left(\mathrm{Q}^1 - \mathrm{Q}^0\right) \mid \mathbf{X}_{\text{FE}} \tag{4}$$

[14] If there is a structural difference between the subject-specific effects of teacher quality the identified effect of (2) would be a weighted average which would change the economic interpretation of β_{DiD}.

where $\mathbf{X}_{\text{FE}}$ comprise teacher year and federal state fixed effects and class fixed effects. This assumption may be credible, as parental background, class, and school effects, and any further individual differences are held fixed. It would be violated, e.g., if the within-class variation in potential test scores is large, which school principals could observe alongside the quality of their teachers. In addition they had to strategically assign teachers (and their quality) to courses and classes such that test scores between courses are, for instance, either compensated or reinforced between subjects. In this case, at least some parts of β_{DiD} in (2) would also capture a selection effect. This, however, is unlikely to dominate the effect, since within classes it appears more plausible that relative advantages in one particular subject cancel each other out. Although I term this strategy differences-in-differences, the argumentation above clarifies the analogy to an instrumental variables approach, where the school principal's assignment is the plausible random assignment mechanism that I exploit for identification.

A1.2 Robustness of the the employed educational expansion rate

As federal state and year fixed effects are used as (the most important) control variables, the main coefficients of interest, β_{FE} and β_{DiD}, are essentially identified by changes from year and federal state-specific means: $d\ln(\#\text{teachers}_{st})$. Instead of using $\ln(\#\text{teachers}_{st})$ as the regressor of interest, one could equivalently have used the residual from the following regression (Frisch-Waugh-Lovell Theorem):

$$\ln(\#teachers_{st}) = \delta_s + \gamma_t + u_{st}$$

, this residual u_{st} equals $d\ln(\#\text{teachers}_{st}) = {}^{d\#\text{teachers}_{st}}/_{\#\text{teachers}_{st}}$, which essentially is the ratio of educational expansion teachers to the projected number of teachers needed in the absence of the educational expansion. Using the notation from Section 3.2, this is p_2/p_1. But as we have seen from Eq. (3), $p_2/(p_1+p_2)$ is considered to be the appropriate leverage by which the average teacher quality of a certain teacher cohort in a federal state is affected by the average quality of the incoming teachers. Thus, does $p_2/(p_1+p_2)$ better capture the quality effects? To check this, one can adjust the residual u_{st} (relative change in the teacher force with respect to the projected number of teachers) to the relative change with respect to all teachers by dividing by $\theta = 1 + p_2/p_1$. [15] Plugging in u_{st}/θ instead of u_{st} in regressions 1 and 2 yields estimates that are presented in Table A1.

[15] The parameter θ can be derived as follows:

$$\begin{aligned} \theta\frac{p_2}{p_1+p_2} &= \frac{p_2}{p_1} \quad \Leftrightarrow \quad \theta p_2 p_1 = p_2(p_1+p_2) \\ \Leftrightarrow \quad \theta &= 1 + \frac{p_2}{p_1} \end{aligned}$$

The results presented in this Table indicate that all specifications are largely insensitive toward whether $^{p2}/_{p_1}$ or $^{p2}/_{(p_1+p_2)}$ are employed in the regressions. Hence, β_{FE} and β_{FE} indeed seem to adjust the effect to the marginal teachers.

Tables

Table A1: Tansformed results

	β_{FE}		β_{DiD}
	Math	Reading	Pooled
	(1)	(2)	(3)
Adjusted measure: $^{u_{st}}/_{\theta}$	-1.372^{***}	-0.708	0.953^{***}
	(0.520)	(0.481)	(0.380)

Notes: Federal-state-by-year-level clustered standard errors in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. This table assesses whether the main effects in Tables 3 and 4 are adjusted appropriately to induced changes on the average "quality" of teachers by incoming educational expansion teachers. The identifying variation plotted in Figure 3 is $^{p_2}/_{p_1}$, but the leverage of educational expansion teachers on the average teacher quality of a cohort of teachers from federal state s in year t is $^{p_2}/_{p_1+p_2}$, as shown in Eq. (3) . Therefore, the identifying variation (the residual from a first-stage regression) is divided by the factor $\theta = 1 + {}^{p_2}/_{p_1}$ and plugged into a second-stage regression.

Table A2: Teacher mobility between federal states

	Number of teachers	Percentage
Teacher does not move	234	73.9
Teacher moves to neighboring states	38	11.8
Teacher moves to non-neighboring states	46	14.3
Total	318	100

Notes: teacher mobility is defined as whether a teacher is employed at a school in a federal state that is different to the federal state in which the teacher graduated from high school.

Table A3: Example structure of the data for the triple differences estimation

Student level				Treatment	Teacher level					
ID	Name	Subject	Test score		ID	Name	Subject	Year teacher turned 19	Federal state	$d\,\ln(\#\text{teacher}_{st})$
1	Alexander Meier	Math	2.34	1	1	Lothar Müller	Math	1980	Bavaria	0.08
1	Alexander Meier	German	1.65	0	1	Lothar Müller	Math	1980	Bavaria	0.08
1	Alexander Meier	German	1.65	1	2	Esther Schulz	German	1978	Hesse	-0.02
1	Alexander Meier	Math	2.34	0	2	Esther Schulz	German	1978	Hesse	-0.02

Notes: This table shows the structure used to estimate the main results of the paper by a triple difference estimation. Her pupil is observed four times – two observations for each subject-specific test score (math and reading) by math (here Lothar Müller, lines 1-2) and German teacher (Esther Schulz, lines 3-4). Within each teacher, the test score outcome of each assigned pupil that relates to the respective subject of the teacher serves as a treatment whereas the other test score serves as the control group. To difference out any subject to account for subject-specific effects, the data are expanded on the pupil level such that treatment and control group are reversed. This expansion of the observations, the standard errors remain clustered at the year the teacher turned 19 and the federal state of the teachers' high school graduation.

Table A4: Number of pupil and number of students used in this analysis and dropping reasons

	(1)	2)	(3)	(4)	(5)	(6)	(7)	(8)
	Plain sample	Higher secondary track students [a]	Western Germany	Gymnasium track	Math or German teachers	Teacher older than 19 in 1970	Math teachers	German teachers
N	87,776	53,745	17,337	10,806	5,683	5,311	2,625	2,855
# Students	24,417	14,532	9,042	5,345	4,259	3,980	2,491	2,620
# Teacher	4,952	2,779	1,206	745	345	322	158	168

Notes:
[a] Students of either Realschule, Gesamtschule or Gymnasium.

Table A5: Descriptives for aspects of teacher's job choice

	Statistics	
	Mean	SD
Reconcilability of job and family	3.353	(0.778)
Possibility to interact with people	3.573	(0.538)
Leisure time	2.125	(0.786)
Salary	2.630	(0.737)
Meet challenges	3.013	(0.669)
Joy to teach	3.691	(0.489)
Job security	3.200	(0.750)
Prestige of being teacher	1.863	(0.776)
Possibility to accomplish things	2.382	(0.759)
Dedication to subject	3.580	(0.553)

Domains of job choice are based on answers on the following question: "How important was the following aspect for your choice of becoming a teacher?" Teachers could respond on a 5-point Likert scale ranging from 1 – "Very unimportant" – to 5 –"Very important".

Table A6: The association between the degree of the relative degree of the educational expansion and the Abitur grade for different samples

	Grade Abitur		
	(1)	(2)	(3)
$\ln(\#\text{teacher}_{st})$	0.942^{***}	0.970^{**}	0.956
	(0.362)	(0.419)	(0.661)
Sample restrictions:			
-Realschule	✓		
-Gymnasium	✓	✓	
-Sample teachers	✓	✓	✓
# teacher	995	625	284

Standard errors in parentheses, $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Each column shows the effect of the educational expansion on the selection of teachers indicated by their grades. The underlying data is on the teacher level. Control variables comprise year fixed effects, federal state fixed effects and subject fixed effects.

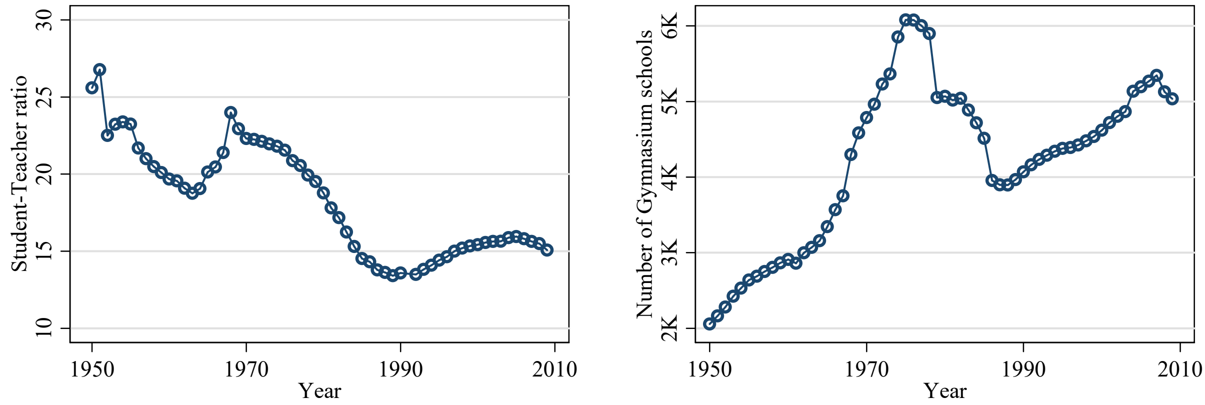
Table A7: Potential impact of teacher non-response on the main effect

	Test scores							
	Math			Reading			Pooled scores	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Valid Teacher	0.138***	0.087***	0.040	0.136***	0.079***	0.032	0.136***	0.137***
	(0.035)	(0.030)	(0.030)	(0.032)	(0.026)	(0.028)	(0.032)	(0.027)
Valid Teacher $\times D$								0.0004
								(0.027)
Cross-subject score[a]		✓			✓			implicitly
School fixed effects			✓			✓		
# observations	15,123	14,621	15,123	14,831	14,621	14,831	29,954	29,954

Notes: This table shows whether information on teacher having a valid interview (a prerequisite for knowing his birth year and the federal state of high school graduation among others) is able to predict the test scores of his students. Columns (1) to (6) report regression coefficients from the fixed effects regression similar to (1) whereas columns (7) and (8) show the results of the difference-in-differences regression as reported in (2). Most importantly, this table shows that while there may be some bias from teacher non-response in the fixed effects strategy (1) (it is still unclear whether this is correlated with the educational expansion), this bias disappears once within-school variation is used (columns (3) and (6)). If the difference-in-differences strategy is employed meaning that within-course variation is solely and effectively exploited, teacher non-response is not at all able to predict the test scores of the students.

[a] The reading score for math teachers and math test score outcomes and math scores for German teachers and reading score outcomes.

Figures



(a) Student-Teacher ratio over time

(b) Number of schools over time

Figure A1: Further characteristics of the educational expansion in Germany

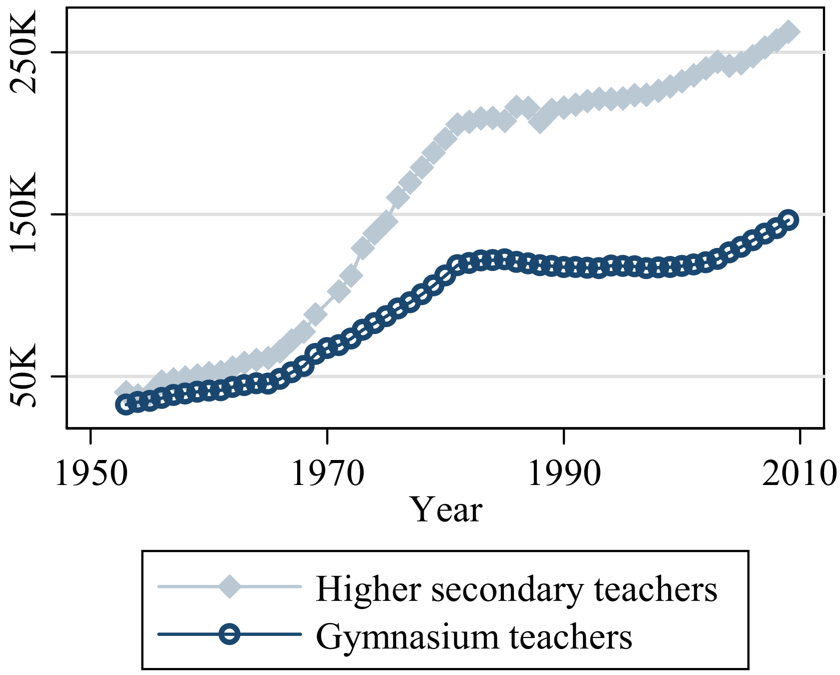


Figure A2: Number of Gymnasium teacher over time

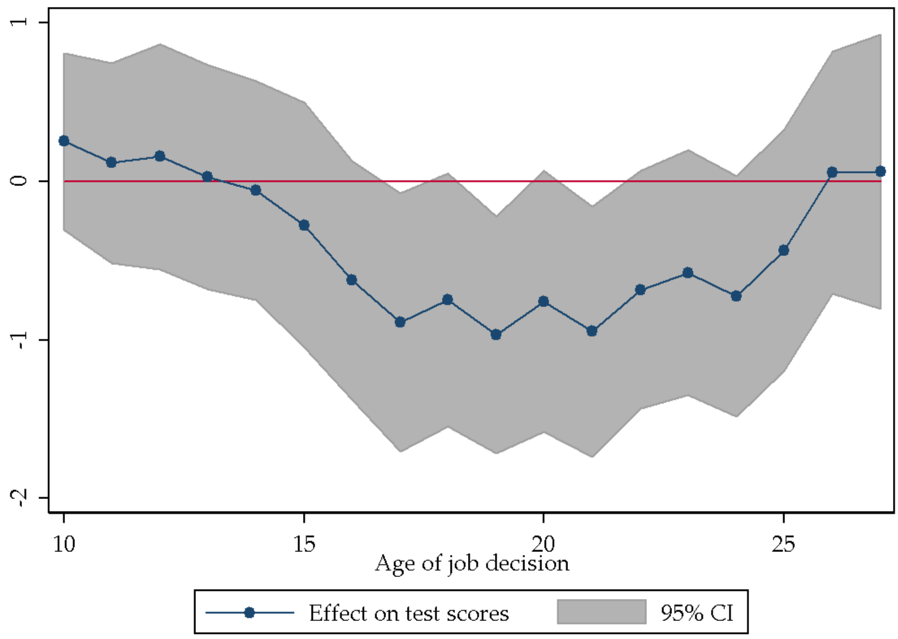


Figure A3: Sensitivity of the effect with respect to the assignment year

This graph plots the effect of EET on student test scores (β_{DiD}) and how this effect changes with respect to a different assignment year. In the main analysis, the assignment year was set to 19, that is, the year of academic track education. The effects are similar in magnitude and precision over the age range from 17 to 21. Outside this range, the effect is negligible.

S1 Supplementary material

Table S.A1: Sensitivity of the results due to selective student non-response and test participation

	(1) Baseline	(2) Class size FE	(3) Class size≥10
$\ln(\#\text{teachers}_{st}) \times D$	-0.966^{**}	-0.961^{**}	-1.0611^{**}
	(0.381)	(0.383)	(0.415)
$\ln(\#\text{teachers}_{st})$	0.335	0.231	0.344
	(0.399)	(0.409)	(0.498)
D	8.384^{**}	8.341^{**}	9.201^{**}
	(3.292)	(3.305)	(3.582)
# observations	10,330	10,330	9,465
# teachers	322	322	251

Notes: The coefficients are estimated using equation (2). Column (1) refers to the results from column (8) in Table 4. Column (2) includes class size fixed effects to see whether the main component of the correlation structure has to be attributed to the class size that may correlate with $\ln(\#\text{teachers}_{st})$. Here, "Class size" refers to the number of students with a valid test score observation. To further see whether the main result is actually driven by very small classes, column (3) drops classes with less than 10 students.

Table S.A2: Potential impact of student non-response on the main effect

	Test score class size	Fraction with valid test score
	(1)	(2)
$\ln(\#\text{teachers}_{st}) \times D$	-0.877	0.014
	(1.761)	(0.112)
$\ln(\#\text{teachers}_{st})$	-13.410^{***}	-0.673^{***}
	(5.955)	(5.369)
# observations	10,330	6,393
# teachers	322	208

Notes: The coefficients are estimated using equation (2). Test score class size refers to the number of valid test score observations by course (math or German). Fraction with valid test score observations divides the the test score class size by the actual number of students per class. Its observation number is lower because of the non-response of the respective class teacher.