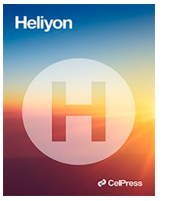
Heliyon 10 (2024) e25043

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Why are there differences across German states in student achievement and cognitive ability?

Heiner Rindermann

Department of Psychology, Chemnitz University of Technology, Wilhelm-Raabe-Str. 43, D-09107, Chemnitz, Germany

ARTICLE INFO

Keywords:
German state differences
Students' cognitive ability
Politics
Intelligence

ABSTRACT

Background: For more than twenty years, large and generally stable differences in academic achievement and cognitive ability have been reported within Germany. In such studies, the southern regions lead in the west and east, while city-states lag behind. Expressed in school learning time, the students in Bavaria are 14 months ahead of the students in Bremen. It is striking that there are no or only marginally received studies on causes and consequences.
Purpose: This study attempts to explore the causes and consequences of the differences within Germany and what can be learned in general about their development.
Materials and methods: We use data from student assessment and other studies (e.g., PISA, IQB) and apply correlational and path analyses, controlled for various background factors.
Results: There are no stable correlations with evolution (genes), educational level of society (adult school years) and wealth (GDP per capita). However, there are high correlations, robust across indicators, with "burgher-conservative" education policies, e.g., central exit examinations, early tracking, grades at a young age (around $r \approx .65$); with measures of students' quantity of education (hours of instruction, no teacher shortage; $r \approx .40$); with measures of tertiary educational quality and appreciation of education (university quality, short duration of studies, professors' salaries; $r \approx .50$); with student native/immigrant ratio ($r \approx .50$); with middle-class burgher lifestyle (less private debt, less welfare dependency and less crime; $r \approx .60$); and with burgher-conservative-right politics (share of votes for CDU/CSU and non-left parties, non-left state governments; $r \approx .80$). Longitudinal analyses over four decades reveal interaction effects, i.e., more burgher policies statistically lead to more cognitively competent students ($\beta \approx .45$) and more cognitively competent populations vote for burgher parties ($\beta \approx .30$).
Conclusions: The results, which support the efficacy a bourgeois-conservative education policy and of lower immigration rates, are delicate for the practice of student achievement research and for the political milieu that dominates the social sciences.

E-mail address: heiner.rindermann@psychologie.tu-chemnitz.de.

https://doi.org/10.1016/j.heliyon.2024.e25043
Received 16 July 2022; Received in revised form 17 January 2024; Accepted 18 January 2024
Available online 24 January 2024

1. Introduction: existence of federal state differences in abilities

The international student achievement studies became known in Germany with the Programme for International Student Assessment 2000 (PISA).[1] Although there were several forerunner studies in the 1960s–1980s, and since 1995 with Trends in International Mathematics and Science Study (TIMSS) also at a high methodological level, it was not until PISA 2000 that results on ability levels for the German states were published. It was leaked from TIMSS 1995 that Bavarian students performed significantly better than those in North Rhine-Westphalia [1], but there were no official results. A study of intelligence testing of draftees in 1992 and 1998, which provided results at the regional and state levels, was published late and has not become widely known [2].

Data and public attention changed with PISA (2000, 2003 and 2006 with data for German states). Later, Progress in International Reading Literacy Study (PIRLS 2001 and 2006) and the within Germany IQB studies (2009, 2011, 2012, 2015, 2016, 2018) were added (acronyms see footnote 1). The mentioned studies differ in the age of the students (PISA 15-year-olds, PIRLS about 10 years, IQB about 10 and 15 years), in their target populations (PISA 15-year-olds, the other 4th or 8/9th grade level), how close scales are to the curriculum (PISA and PIRLS are less close to the curriculum, IQB is close to the curriculum), in the chosen dimensions (reading German, orthography, foreign languages, mathematics, natural sciences), somewhat in the definition of migration background and in the additional scales used (personality, behavior at school and at home, teaching methods, school, parental background).

Ability differences within countries at the state or regional level have been studied many times, for example, for the U.S., the UK, Spain, and Italy [3–6]. However, there are hardly any studies on the causes and consequences of the differences within Germany. Why do students get better results in the south than in the north? Do the differences in ability show consequences for the development of states as they have been demonstrated for countries at the international level [7–9]?

1.1. Political factors: heterogeneous findings or a moderate burger-rational view?

Given the good data situation, it is surprising that the (German) scientific community has not addressed the differences between the federal states within Germany. There is a publication by the Finnish scientist Eka Roivainen from 2012 [10] on the topic, but according to Scopus it has never been received or cited in German-speaking countries. A 2003 presentation by Rindermann at the Educational Psychology Conference in Bielefeld, Germany, in which he developed the idea that conservative educational policies emphasize the aspect of achievement, effort, and high standards, thereby promoting ability development, was well attended at the time but failed to stimulate further research. In 2019, a similar talk by the same author in Leipzig at the same conference was attended by only a few people.

The possible connection between student test scores and political factors was subsequently discussed only in a journalistic and political context, e.g., by the then school principal and president of the German Teachers' Association Josef Kraus [11]:

> "In general, the so-called A-states (SPD-governed) perform considerably weaker than the so-called B-states (Union-governed)."[2]

The findings have been an open secret ("mokita") in Germany for decades, as journalist Heike Schmoll wrote, for example [12]:

> "As early as 1999, the SPD Working Group for Education intervened in a protest letter to the social democratic school ministers against the planned intra-German PISA comparison. One feared that the education policy of one's own party was doomed to failure. 'It can be predicted without testing that states with selective school systems that have resisted the school structure reforms of the last 30 years will have better student performance in all types of schools,' the letter said. The GEW[3] echoed the warning at the time, seeing the same result coming, stating that 'anything else would be a sensation.'"

However, it is by no means clear from international research that conservative attitudes and policies are associated with higher cognitive ability or achievement. On the contrary, most studies show an affinity between progressive beliefs and intelligence: Conservatism on the left-right scale is associated with lower ability in international comparisons, e.g., using data from the World Values Survey ($r = -.25$, at the level of $N = 92$ countries).[4] Especially traditionalism is associated with lower cognitive ability in international comparisons, e.g., the more people in a country believe in god the lower is the average intelligence ($r = -.66$, $N = 79$ countries).[5] This pattern is supported by a multilevel study of Stankov from 2009 [13] using various indicators of religious attitudes, general values, and personality traits that showed negative correlations with intelligence (individual level: $r = -.28$, country level: $r =$

[1] Student achievement studies or educational assessment studies: *PISA*: Programme for International Student Assessment, since 2000 every three years, measures reading literacy, mathematics and sciences of 15-year-old students in secondary school. *TIMSS*: Trends in International Mathematics and Science Study tests 4th and 8th graders in math and science, since 1995 every four years. *PIRLS*: Progress in International Reading Literacy Study, since 2001 every five years, measures reading literacy of students in grade 4 in primary school. *IQB*: Institut zur Qualitätsentwicklung im Bildungswesen, German student assessment studies 2009, 2011, 2012, 2015, 2016 etc., measure curriculum-related reading in German, orthography, foreign languages, mathematics and sciences of students in grade 4 and 9 in primary and secondary school. The IQB is a broader and more curriculum-based achievement test.

[2] Translated by the author, as are all other quotations originally written in German in this paper.

[3] GEW: Gewerkschaft Erziehung und Wissenschaft, Education and Science Workers' Union, a traditionally left-wing trade union.

[4] World Values Survey (2014), data from 1981 to 2014, variable e033: "Self positioning in political scale", 1, left, to 10, right; cognitive ability data based on student assessment and intelligence test data, country level data (Rindermann, 2018, updated March 2022). [21]

[5] World Values Survey (2014), variable f050: "Belief in God", 0, no, 1, yes. [21]

−.73). Other studies came to similar confirming findings [14,15]. But the results are not so simple. There are many opposing results, e.g., Carl found at the individual level in the United States and at the regional data level in the United Kingdom that intelligence is associated with more conservative political beliefs [16,17].

Rather unusual is the result of a study that showed that intelligence is not associated with left or right, but with moderate, middle-of-the-road political orientation, here in a Brazilian sample [18]. The tendency was for moderate-reasonable positions to be associated with higher intelligence rather than extreme ones. Political positions should also be understood in a more differentiated way, usually distinguishing between an economic dimension for or against high government spending and more welfare state and a socio-cultural dimension for or against abortion, adoption rights for homosexuals [19]. Intelligence seems to be supportive for a more economist view, i.e. less government spending ("intelligence makes people think like economists" [20]), and for a more modern social-cultural view [14]. Presumably, it is not so much conservatism or progressivism per se that is associated with intelligence, but specific traits in these that are promoted in the direction of more rationality through intelligence. Such a position is supported by a study by Ganzach from 2018 [22], who showed that within given beliefs, intelligence leads to a more consistent-rational form of belief.

Depending on the country and historical time, more left-wing or more right-wing positions may be more reasonable and thus associated with intelligence. Intelligence helps to recognize reality and problems, to understand complex tasks and to come to a better solution by weighing side effects and advantages and disadvantages on different sides [23]. Of course, there will never be complete agreement on what is the rational solution to a concrete problem in a concrete situation (see e.g., Pinker, 2021 [24], and the discussion around his book), and there can probably be different solutions in the reality of complex life problems, but intelligence is a crucial aid in arriving at more rational results.

What has not been discussed at all in all these studies so far is that not only can cognitive abilities have an effect on political orientations, but also vice versa, there can be an effect from political orientations on cognitive abilities. Attitudes in general, including political orientations, change society and institutions and influence daily life, e.g. what money is spent on, how schools and lessons are designed and how children are educated [25,26]. For example, Becker and Woessmann in 2009 [27] have shown that Protestantism has a positive impact on education, as does Christianity in Africa [28]. Of course, there is no direct impact, beliefs do not create abilities, but beliefs change behavior and the environment, which can influence the development of abilities.

Applied to the German situation, for decades burgher-conservative-right politics went hand in hand with more traditional, strict, and achievement-oriented school policies, such as a focus on objective central exams, high standards, and discipline. Over decades, these educational policies, if supportive, may have ultimately led to improved student achievement.

1.2. Further factors: geography?

While a north-south gradient can be observed at the international level (precisely: the measured cognitive abilities increase on average with the distance from the equator, $r = .69$; Rindermann, 2018, p. 286 [29]), within Germany it is exactly the opposite: there is a south-north gradient (Fig. 1). However, geography is only descriptive, not a causal theory, and metereological and similar differences are far too small in Germany to have an impact.

1.3. Education

Many educational, school, and instructional characteristics that have been positively associated with ability in individual- and class-level studies and for which there is elaborate theoretical support for causal effects have also been positively associated with ability at the country level. For example, the more adults have a secondary school degree, the longer they have attended school, the better are on average the results in student assessments and intelligence tests ($r = .74$; p. 266 [29]). Other positively associated variables include kindergarten attendance, discipline, and teacher quality. All these characteristics represent quantity and quality of education that support the development of ability. If differences between German states are found in such educational variables, then they could help explain state differences in tests results.

1.4. Economic conditions: wealth (income), debts and unemployment

At the international level, cognitive ability has a positive impact on income, e.g., as measured by GDP per capita, which has been confirmed in many studies [7–9,30]). At the individual level, there are similar positive effects of intelligence on income (in Germany by Kramer in 2009 [31]; in Britain by Irwing & Lynn in 2006 [32]) and more generally on job training and job performance (in Germany by Hülsheger et al. in 2007 [33]; in the U.S. by Schmidt & Hunter in 2004 [34]). Higher intelligence enables better and faster learning, more complex tasks can be mastered, innovations are developed etc., all together leading to more income. Higher debt may be the result of lower income, but it may also stem from problems managing the given income and making spending choices. This is also a cognitive task.

However, there may also be reciprocal effects: Better living conditions lead to better health, less stress, more stimulation, etc., thereby promoting cognitive development. The authors of the German Armed Forces Study, who found differences in intelligence across states, explicitly assumed that unemployment lowers intelligence [2,35]. They presented their conviction already in the subtitles of their publications: "Effects of unemployment and intra-German migration" and "Loss of human capital in regions of high unemployment". Unemployment can lead to paralysis, hopelessness and apathy, which are passed on to children through the family [36]. The more highly qualified migrate from regions with higher unemployment.

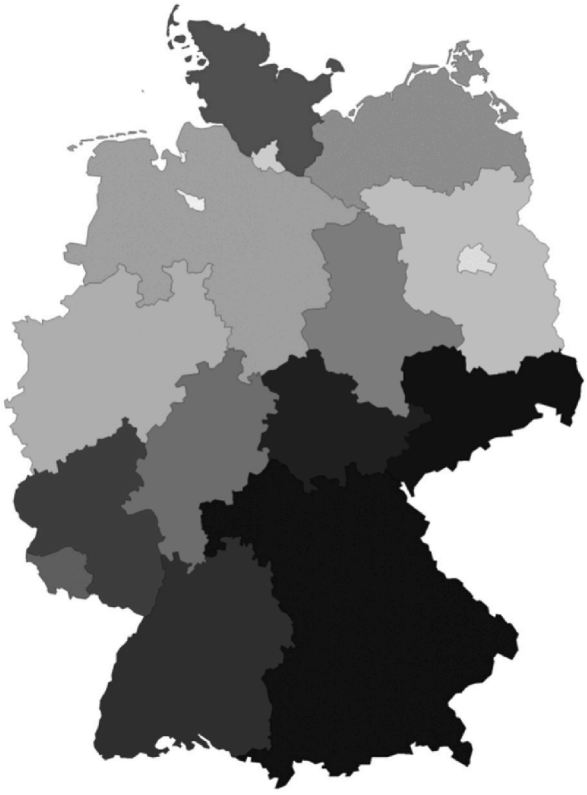


Fig. 1. Mean value in studies of cognitive abilities for 16 German states (1998–2016) (darker = higher and better results).

1.5. Demography

Another frequently discussed factor that may explain differences in test scores within Germany is the proportion of pupils with an immigrant background per state. The educational researcher Klaus Klemm, who was a member of the scientific advisory board of the German PISA studies, and Petra Stanat, the head of the IQB, have mentioned this factor several times [37–40]. However, at the international level, there is no negative correlation between the proportion of immigrants and test scores in cognitive ability test results ($r = .07$; p. 266 [29]). Immigrants vary greatly by region of origin and country of destination. The only thing that was robust was that migrants are mainly attracted to wealthy countries that also tend to score well on tests, i.e., where the population has high levels of cognitive ability. Systematic analyses for Germany and the differences between the federal states are lacking.

1.6. Background factors: genes and culture

At the international level, indicators of the two background factors of genes (evolution) and culture (religion) correlate highly with student assessment scores and psychometric intelligence tests. At the individual level, the evidence for intelligence-coding genes and their mechanism via proteins to neurological structures and cognitive processes and ultimately to intelligence is rather limited, but there is a lot of indirect evidence from heritability and genome-wide association studies (GWAS) [41,42]. The same applies to student achievement in schools [43]. Additionally, genes found in GWAS at the individual level are also associated with ability differences at the country level [44]. Indirect evidence also comes from studies on haplogroups: Haplogroups are not intelligence-coding genes, but are indicators of evolutionary processes and ancestry. Such haplogroups correlated highly in an international sample with ability levels ($r > |.80|$ [45]).

In terms of culture, Protestantism has always shown a positive effect on education and cognitive ability in German, European, and global analyses (compared to Catholicism or in larger samples of different religions [27–29,46]). Of course, religion does not work directly, a baptism does not make one smart, but religion can work through a change in attitude and then behavior, e.g., reading more [47].

In summary, there is much to suggest that both background factors could be at least partly responsible for the state differences within Germany, which is why it is important to test them.

1.7. Indirect and reciprocal effects, ecological fallacy

Until now, we have only talked about single factors. However, it is unlikely that factors will act independently of each other. Rather, it is likely that they operate together and indirectly, such as cultural or political beliefs via educational institutions and behavior. Finally, reverse effects may also occur, i.e., not only do policies affect ability, but ability affects political attitudes, as tends to be assumed in the intelligence literature [14,48].

While these assumptions are theoretically compelling, it is difficult to translate them into appropriate statistical designs, especially in a sample with only 16 states. We attempt to address the problem by a) using different indicators for each construct and testing the stability of the findings across these different indicators, b) constructing scales with higher reliability and validity than single items, c) testing the stability of the findings across different methods (regressions, path models), and d) averaging in longitudinal analyses the results across different time periods (to obtain values that compensate for deviations over decades). The problem of the ecological fallacy is addressed by comparing findings with those at other levels (individual, class, school, countries [49]) and by theoretical considerations regarding possible mechanisms of effect. At the end, we repeat the analysis with election results at the level of 83 districts.

1.8. Outline

The paper continues as follows: In the Methods section (Appendix), we describe the origin of the data, starting with competency studies, then possible independent and dependent variables, and finally our statistical analysis methods. The Results part begins with descriptive data, the state means in various modern and older competence studies, and then continues with correlations with possible causal factors. The Results section ends with combined analyses of theoretically and empirically important factors, including the analysis of interactions. In the Discussion section, we highlight the relevant results and interpret them; in the Limitations, we primarily address the problems, especially of the small sample. Finally, we ask why so few analyses are conducted on the differences between the German states and summarize the key findings in Conclusions.

2. Research questions

The following questions will be addressed:

1. Is it possible to assume a state overall score in cognitive ability or are there substantial differences between studies, dimensions, student groups, ages, grades, and measurement times?
2. Are there correlations with possible causal factors? Can possible causes (and consequences) of the German state differences in ability be proven theoretically and empirically?

3. Method

We almost exclusively use data provided by other scientists and institutions. In this way we reduce possible biases caused by our own assumptions. The data sources (i.e., student achievement and cognitive ability studies, educational factors, aspects of tertiary education and research, of economy and society, of religion and culture, of politics) are listed in detail in the online Appendix. As statistical analysis methods we used means, correlations, regressions, cross-sectional path models, and longitudinal path models comparing cross-lagged effects (see online Appendix).

4. Results

4.1. Means and magnitude of ability differences

Table A1 shows the mean results in student achievement and cognitive ability studies between 1998 and 2016 from four different sources (PISA, PIRLS, IQB, German army examination test). Students in Bavaria, Baden-Württemberg, Saxony and Thuringia show the best results, while Bremen, Berlin, Hamburg, Brandenburg and North Rhine-Westphalia bring up the rear. The size of the state

Table 1
Correlations among ability studies (means across scales and years).

	Student mean	PISA mean	PIRLS mean	IQB mean	Army mean
CA mean	.99	.94	.89	.94	.68
Student mean		.92	.93	.96	.57
PISA mean			.75	.84	.71
PIRLS mean				.86	.39
IQB mean					.51

Notes. Example: states that perform well in PISA also perform well in IQB studies ($r = .84$). $N = 16$ German states. Within studies across time, PISA 2000 to 2006: mean $r = .90$; PIRLS 2001 and 2006: $r = .76$; IQB from 2009 to 2016: mean $r = .74$. Grand mean: $r = .80$.

differences is substantial: assuming that 42 SASQ points are gained per school year, students in Bremen (student achievement mean 465, IQ 95, corresponds to an ability level of 8;10 [years;months]) lag Bavarian students (SASQ 516, IQ 102, 10;0) by one school year and two months (SASQ 50.48, IQ 7.58). Fig. 1 shows the distribution of the overall score in a map.

4.2. Correlations within and between ability studies and across time

Can we use a general overall state ability value? Table 1 shows the correlations between the means within and of different studies (e.g., "PISA" is the mean of the three PISA scales and the three measurement points 2000, 2003 and 2006). The results are robust across different studies, dimensions, student groups, ages, and measurement times (see Tables 1–5 and Fig. 2). Correlations between study means are high (Table 1). Among student achievement studies they are higher than with the (one) 1998 Armed Forces Intelligence Test Study, which also includes only males and among them no conscientious objectors, and it is also the oldest survey. The homogeneity resp. reliability of the overall student achievement score (PISA, PIRLS, IQB) is Cronbach-$\alpha = .93$, with the army study $\alpha = .89$. In a factor analysis, the first unrotated factor explains 76 % of the variance, and all four studies load at least with $\lambda = .72$ on this factor – so there is a strong "big" *G*-factor at the state level.

States with good student performance in PISA reading comprehension also have good performance in mathematics and science (Table 2, both $r = .94$). The correlation between mathematics and science is close to 1 ($r = .98$).

There are high correlations between the IQB dimensions German, mathematics and science (Table 3), but not with results in English as a foreign language (not added to the overall IQB score). In West Germany, better scores are achieved in English than in East Germany (IQB mean English: $M_{West} = 496$, $M_{East} = 480$). West German students are about equally well in all dimensions, while East German students are weaker in English and better in science (IQB mean science: $M_{West} = 496$, $M_{East} = 529$). One cause may be the education of teachers and thus the quality of teaching: Fewer English teachers in East Germany have studied English (except for Saxony, between 58 % and 67 %) than in West Germany (between 82 and 100 %; p. 101 [50]). Taking out the west-east difference by partial correlations, the correlations between IQB dimensions increase and all are positive (in brackets in Table 3).

In German states where students without a migration background ("natives") show better student achievement test results, students with a migration background ("migrants"; Table 4) also show better student achievement. This speaks for a general quality of education that is effective. However, the correlation is not high at $r = .45$ and hardly increases when the west-east difference is removed ($r_p = .48$). One reason for this could be that the variable migration background is at least somewhat error-prone: many pupils, and especially those in countries and states with weak results, do not indicate whether they have a migration background, and this group performs particularly poorly. Probably more importantly, the origin and average ability level of immigrants may vary from state to state. For example, migrants come from Eastern Europe (e.g., Poland or Ukraine; with average to slightly below average results), East Asia (e.g., from Vietnam with good results [51]) or from the Middle East and West Asia (with below average results). However, this was not systematically reported for each state by the student achievement studies.

The older ability test studies were not considered for the overall average. They serve to check the stability of the ability patterns over decades. These older data have problems: First, for all data prior to 1990, the number of states is reduced to 11 – no information is available from Eastern Germany. Then the data quality is significantly worse than in modern studies. Flor, Ingenkamp, and Schreiber, for example, warned against unrepresentative surveys; moreover, teachers themselves have administered the tests [53]. Nevertheless, positive correlations are found with results of modern studies (PISA etc.), on average $r = .44$ (Table 5, mean of all correlations). The 1992 German army test of male conscripts isan outlier. This is the first measurement to include data from East Germany. However, when we partial out west-east and city-rural state differences the correlations with the 1992 German army tests become very high ($r_p =$ .86, .82 and .70). At that time, young adults in West and East Germany looked back on very different developmental conditions in their youth. East Germans caught up strongly between the fall of the Berlin Wall (1989) and the year 2000, here in intelligence, but also in height [56]. In 1992, the difference in IQ was about 5.74 IQ points, in 1998 about 1.49 IQ points (based on male conscripts). Taking the *g* factor of all six older achievement and ability studies (last row of Table 5) results in correlations of about $r = .63$. The cognitive ability pattern of the German federal states has been reasonably stable for decades.

4.3. Correlations, possible causes and consequences: education

At the international level, adult educational attainment and cognitive ability are highly correlated ($r = .74$; p. 266 [29]). At the individual level, one year of schooling boosts intelligence by 3.39 IQ points [57] (Germany, in some cases even more [58]). In turn, more intelligent persons go to educational institutions longer. Students at the German "Gymnasium" (grammar school, from grade 5 to 12/13) have an IQ about 30 IQ points higher than those from the "Hauptschule" (lower secondary education school at level 2, from grade 5 to 9/10 [59]) and go to school two or three years longer, and most Gymnasium students also go on to study at university. There is a positive interaction between educational attainment and ability, but probably with diminishing returns. At the international level, similar interaction processes are to be expected, e.g., that better-educated adults promote education in a country, both privately and by influencing politics and society. Within Germany among the 16 states, however, there is (uncorrected) no such positive correlation ($r = .01$; Table 6). Nevertheless, the results of the partial correlations are consistent with the hypothesis ($r_p = .46$). The reason for this discrepancy is that the level of adult education is higher in the city states ($d = +.94$), but student ability is lower there at the same time ($d = -1.81$, *SD* at the state level). Cities attract well-educated adults, but also less educated immigrants who, on average, have more children.

The first four variables of Table 6 (adult educational level, kindergarten attendance, amount of instruction, no teacher shortage) stand for *amount of education*. The tendency is that more education is associated with higher cognitive ability (with mean scale: $r =$

Table 2
Correlations among PISA dimensions.

	Mathematics	Science
Reading	.94	.95
Mathematics		.98

Notes. Means based on 2000, 2003 and 2006 surveys; "Mathematics" is the mean of the three measurements in 2000, 2003 and 2006; $N = 16$ German states.

Table 3
Correlations among IQB dimensions.

	English	Mathematics	Science
German	.49 [r_p .78]	.91 [r_p .97]	.53 [r_p .72]
English		.18 [r_p .73]	−.35 [r_p .40]
Mathematics			.78 [r_p .75]

Notes. 16 states; IQB (German student assessment study) surveys 2009, 2011, 2012, 2015, and 2016 combined; "German" etc. is the mean across different measurement points; correlations in square brackets [r_p] represent partial correlations, west-east difference removed.

Table 4
Correlation between natives' and immigrants' ability levels.

	Migrants' mean
Natives' mean	.45 [.48]

Notes. $N = 16$ states; based on PISA, PIRLS and IQB; natives: students without a migration background, migrants: students with a migration background; correlation in square brackets is partial correlation, west-east difference removed.

Table 5
Correlations with older cognitive achievement and ability studies.

	Cognitive ability mean	Student mean	Grammar school mean	N
Reading in grammar school (Gymnasium) 1969	.33	.32	.26	11
Student achievement 2nd grade 71 Flor	.56	.53	.50	10
High-school diploma test (Abitur) 1973 Trost	.67	.71	.76	11
Student achievement AST 4th grade 1991 Flor	.49	.57	.36	15
Army mean 1992	.10	−.03	.19	16
Preliminary medical examination 1994–2000	.34	.24	.53	14
g factor of older cognitive ability studies	.63	.63	.63	16

Notes. Reading in Gymnasium 1969 documented by Baumert et al. (2003, p. 321 [52]); 71 Flor: results in achievement tests RT2 and AST2 (arithmetic and German) in primary school 2nd grade in 1971 (Flor et al., 1992 [53]); "Abitur" 1973 Trost: German and mathematics test of high school graduates in 1973 (Trost et al., 1976 [54]); AST: General Student achievement Test for primary school 4th graders in 1991 (Flor et al., 1992 [53]); Army mean 1992: male conscripts in an intelligence test (Ebenrett et al., 2002 [35]); Preliminary medical examination 1994–2000: Results of the first examination in the study of medicine 1994 to 2000 at the university (Physikum; Robra & Schmitt, 2001 [55]). Number of states in last column.

.61), more brings more, this in line with many studies. The other variables (from Central exit examinations to Achievement structure) stand for a more *conservative*, traditional, burgher, achievement-oriented or for a more *progressive* equity-oriented education.[6] In Germany, the more conservative parties and states have always been skeptical about an expansion of the Gymnasium (low high-school rate, "Abitur" [60]; $r = -.58$ and $-.81$ with conservative orientation of state governments, Table 12). However, from the amount of education view, there is a paradoxical result: the more young people go to the Gymnasium and achieve a high school diploma, i.e. go longer to more demanding schools with classmates at higher cognitive levels, the lower the student achievement test results are on average in such a federal state ($r = -.75$). This contradicts the usual positive effects of longer schooling. Within the education paradigm, this can only be explained by lowered standards and other detrimental changes in schools (e.g., less discipline) that may accompany such progressive-liberal education policies (see *Limitations*).

[6] SPD-governed countries, for example, were long against a central high school diploma (Central exit examinations), while CDU-governed countries were more in favor of it.

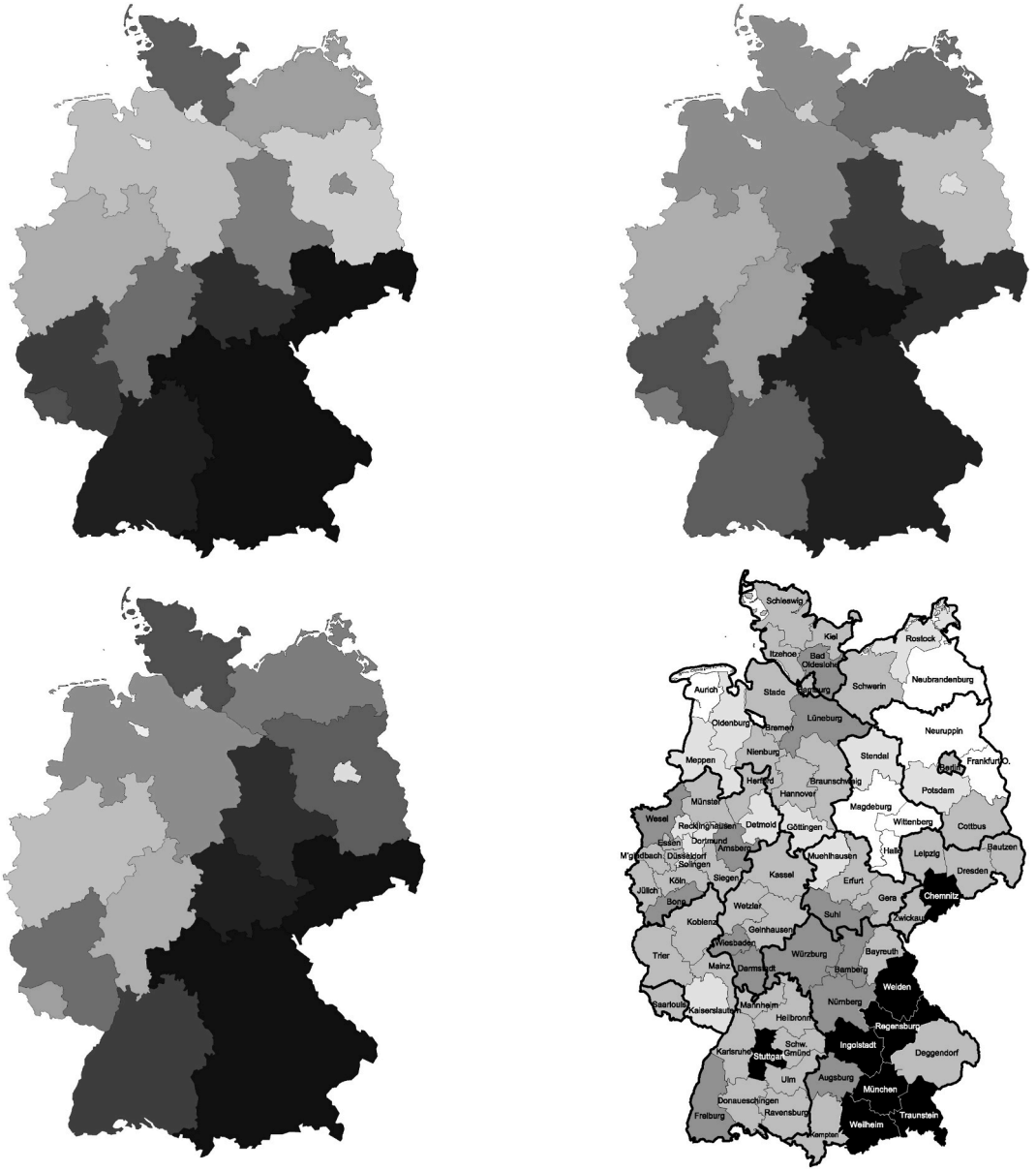


Fig. 2. Mean of three PISA-studies 00 03 06 (top left), of two PIRLS-studies 01 06 (top right), of five IQB-studies 09 11 12 15 16 (bottom left) and one German army-study 98 (bottom right; from Ebenrett et al., 2002, p. 11, reprinted with permission [35]) (darker = higher).

As is the case internationally (e.g., $r = .11$; p. 266 [29]), *centralized final exams* are positively associated with cognitive competence. Centralized exams provide a more reliable measure of competence, they define a minimum standard, signal that achievement is important, motivate learning, improve discipline, and their results inform students, parents, teachers, and school administrators [61]. This is also found in our within Germany comparison ($r = .76$). All correlations with the different ability variables, across different ability levels, and excluding differences between West and East and between urban and rural areas (partial correlations) are stable, i.e. they are independent of concrete operationalizations, performance levels and regional conditions. Our results are supported by an older econometric study by Büchel et al. from 2003 [62], which used another method (differences-in-differences) and found a similar positive effect for Germany.

Tracking at a young age allows for better alignment of instruction with student ability levels, signals that achievement is important (even in the period before tracking, i.e. in elementary school), and increases the average class performance of better students, which again boosts instruction and learning [63]. As in country comparisons ($r = .28$; p. 266 [29]) positive correlations are found within Germany ($r = .36$). The small negative correlation with the ability level of gifted students is strange ($r = -.15$). Usually the gifted benefit more from tracking, in studies at the individual level: $d = .12$ in general but $d = .30$ for better students [64]; and in a

Table 6
Education.

	Cognitive ability mean	Student mean	95 %-level	05 %-level
Indicators of amount of education (quantitative)				
Adult educational level	.01 [.46]	−.05 [.39]	.49 [.74]	−.16 [.24]
Kindergarten attendance	.30 [.16]	.39 [.18]	.43 [.25]	.45 [.05]
Amount of instruction (hours/year)	.77 [.70]	.77 [.69]	.35 [.18]	.76 [.73]
No teacher shortage	.51 [.61]	.59 [.62]	.59 [.56]	.62 [.48]
Quantity of education mean	.61 [.83]	.65 [.80]	.71 [.73]	.64 [.66]
Indicators educational policy progressive-equity-left vs. conservative-burgher-right (political)				
Central exit examinations (1995)	.76 [.80]	.79 [.84]	.63 [.53]	.72 [.73]
Tracking at a young age	.36 [.22]	.35 [.24]	−.15 [−.26]	.32 [.24]
Discipline	.71 [.43]	.77 [.45]	.81 [.90]	.82 [.43]
Comprehensive school rate	−.74 [−.56]	−.76 [−.61]	−.26 [−.06]	−.70 [−.52]
Gymnasium rate	−.45 [−.22]	−.42 [−.28]	−.05 [−.02]	−.38 [−.34]
High-school rate (Abitur) (1995–2000)	−.75 [−.49]	−.75 [−.54]	−.39 [−.31]	−.71 [−.48]
Achievement requirements good & high	.88 [.87]	.92 [.94]	.79 [.73]	.84 [.83]
Traditional instruction (PISA)	.55 [.46]	.63 [.54]	.42 [.19]	.62 [.38]
Grading in primary school (at age 7/8)	.40 [.01]	.44 [.02]	.46 [.29]	.46 [−.05]
Amount of homework in grade 4	.25 [−.17]	.28 [−.23]	.19 [−.11]	.47 [.02]
Klafki proximity	−.83 [−.78]	−.82 [−.81]	−.65 [−.67]	−.76 [−.80]
Achievement structure	.75 [.63]	.78 [.67]	.48 [.33]	.71 [.54]
Conservative educational policies mean	.90 [.77]	.93 [.83]	.56 [.45]	.90 [.75]

Notes. Variables see method section. 95 %-level: student ability level at 95 % threshold, intellectual classes; 05 %-level: student ability level at 05 % threshold, low ability. Partial correlations in square brackets (west-east and city-rural state differences partialled out); except for discipline and achievement requirements ($N = 14$), always $N = 16$ states.

cross-country comparison: $r = .26$ (05 %-level) and $r = .35$ (95 %-level) [29]. We have to bear in mind that the sample in our analysis only consists of 16 states and that single outliers or local peculiarities can easily bias the results. We always have to look at the big picture, across different variables and with different ability levels. We used a scatter plot to test whether an outlier was responsible for the implausible negative correlation. We removed Bremen (tracking at age 10 and lowest 95 %-level result), but the correlation remained negative (95 %-level, $r = -.12$, $r_p = -.11$).

Discipline (as internationally; $r = .49$; p. 266 [29]) is a robust and positive factor in successful teaching ($r = .71$; Table 6). Discipline and effective class management increases learning time, time on task [64].

(Lower) *Comprehensive school,* (lower) *Gymnasium* and (lower) *high school ratios* represent a more conservative orientation of education (see above). The correlations in the within Germany comparison are negative ($r = -.74$, $-.45$ and $-.75$). In German states where more students attend comprehensive schools and more attain the "Abitur" the average abilities are considerably lower. This cannot be a direct effect (comprehensive school is not necessarily worse than lower secondary education/Hauptschule and attending the Gymnasium definitely means more time at demanding schools), but an indirect one, because other characteristics are associated with it, such as a stricter achievement structure.

High achievement requirements, *early grading* and *achievement structure* (more guidelines regarding structure, content and performance assessments) stand for an emphasis on the performance aspect in school. The lead of East Asian countries in student assessment studies, for example, is explained by this emphasis [65]. In our German sample, the correlations are positive ($r = .88$, .40 and .75). *Homework* is positively associated with test results ($r = .25$), but not stable. A global measure of educational policy is "*Klafki proximity*" standing for progressive politics. It is strongly negatively correlated with ability across all ability measures ($r = -.83$).

Finally, more *traditional instruction* stands for more direct instruction, which is one of the most effective instructional methods (individual and class level studies [64]). The mean scale of burgher-conservative-right education policies shows a high correlation with the cognitive ability mean ($r = .90$).

Comparing the results for the intellectual classes and the students at the lower ability threshold (95 % and 05 % levels), it seems that the ability level of the lower achievers depends more on educational conditions or, more generally, on environmental conditions. The average correlation across scales at the 95 % level is $r_{M95\ \%} = .46$, at the 05 % level it is $r_{M05\ \%} = .61$ (averaged using Fisher's *z*-transformation, correlations with comprehensive school, Gymnasium, high-school rate and Klafki proximity inverted). It is possible that the most intelligent students are more able to detach themselves from negative (or positive) environmental conditions; they could be more internally controlled in their competence development.

4.4. Correlations, possible causes and consequences: higher education and science

Higher education and science can be both a determinant of ability (improving the quality and quantity of education, e.g., via teacher education) and a consequence of ability (smarter people go to university and improve institutional quality and science).

There are substantial positive correlations with tertiary education (Table 7): the higher the level of cognitive ability, the better the universities in a state, the faster students study, the higher the professorial salary, and the more salary has been added in the last decade. All correlations are positive, around $r = .30$ to .70. The measures of short duration of studies, from different decades, were uncorrelated with each other, but both correlated positively with ability ($r = .85$ and $r = .18$). Accounting for GDP per capita even

slightly increases the correlations of salary with ability (from $r_p = .55$ to $r_p = .58$ and from $r_p = .33$ to $r_p = .51$; first column). The ability level of intellectual elites seems (plausibly) more relevant to the quality of higher education ($r_{95\ \%} = .40$ vs. $r_{05\ \%} = .06$) and for professors' salary (all six correlations averaged: $r_{95\ \%} = .38$ vs. $r_{05\ \%} = .28$). This pattern is also found for various indicators of cognitive high achievement at the international level (Rindermann, 2018, p. 122, Table 4.6, [29,66,67]. Understanding this pattern as an effect of high ability is theoretically convincing and empirically supported.

Interesting is the considerably high correlation of the ability level of the low ability groups with short studying time ($r_{95\ \%} = .43$ vs. $r_{05\ \%} = .70$; Table 7). Most probable those people at the 05 % ability level do not go to university, but it may suggest that long periods of study are more of a problem for students with lower ability. The high-ability students always have a sufficient level of competence, even if it is lower than in other states, to complete their studies in the given time frame (e.g., at 95 % Bavaria IQ 124 and Bremen IQ 119; see Tables A2b and A2a in the Appendix).

4.5. Correlations, possible causes and consequences: economy and society

Greater wealth (GDP per capita) is not robustly positively associated with ability at the state level (Table 8). The correlations jump from negative to positive when we remove west-east and city-state differences ($r = -.33$ vs. $r_p = .31$). Coming out of the communist GDR period, the Eastern German states are still poorer regardless of cognitive capital, while the urban states are richer regardless. However, all *debt* and *poverty* indicators are robustly and highly negatively correlated with ability (e.g., welfare receipt: $r = -.83$ and $r_p = -.70$). Also *crime* is negatively associated with ability ($r = -.77$ and $r_p = -.44$).

And always, the indicators of problematic societal development are plausibly more highly associated with the lower than the higher ability level (e.g., welfare receipt $r_{95\ \%} = -.55$ vs. $r_{05\ \%} = -.73$, crime $r_{95\ \%} = -.36$ vs. $r_{05\ \%} = -.81$). This suggests that higher cognitive human capital levels are responsible for more successful lives – being able to support oneself, not to get on the wrong track – especially in the low cognitive ability groups. Also, the *social gradient* of how closely parent-SES and student ability scores are related in PISA is highly negatively and robustly correlated with ability and more strongly correlated with the 05 % level than with the 95 % level ($r_{95\ \%} = -.46$ and $r_{05\ \%} = -.87$).

4.6. Correlations, possible causes and consequences: demography (immigration)

Assuming that students understand the language of instruction well, origin (migration background) as such is irrelevant to student achievement, intelligence, and knowledge. And several immigrant groups perform robustly well: East Asians in the U.S. do better in school than (European-ancestry) Americans, Vietnamese in Germany do better than Germans, (European) Jews do better than Palestinians in Israel-Palestine, and so on. In Western Europe, however, a migrant background is usually associated with weaker student performance [68,69]. Differences between German states are also negatively correlated with the share of students with an immigrant background (cognitive ability mean; $r = -.51$, Table 9). The more native students in a state, the better the student achievement in general (student achievement mean; $r = .60$), especially the student achievement of the weakest students ($r_{95\ \%} = .39$ vs. $r_{05\ \%} = .73$) and especially of immigrants themselves ($r_{Mg} = +.42$). The more natives in a state (i.e., in school and classroom) the better are the test results of immigrants: $r_{Mg} = +.42$; there is no positive effect for natives: $r_{Na} = -.13$. The patterns are not always robust when west-east and city-rural state differences are removed. There are many more children with migrant backgrounds in the west ($r = .84$ with west-east) and in urban states ($r = .69$ with city-rural). With a sample of only 16 states and highly correlated characteristics, such variation of results is to be expected.

Migrants from educationally disadvantaged (low educated) families lead to poorer student achievement results and impair, e.g., slow down, teaching. This is particularly serious for the migrants themselves ($r_{Mg} = -.42$) and for weaker students in general ($r_{05\ \%} = -.73$). Studies by the OECD from 2018 [70] also show that the ability development of native students is impaired by a large number of (less educated) migrants. However, the percentage of learners with a migrant background is not decisive for the state differences.

4.7. Correlations, possible causes and consequences: culture

At the international level, differences in education and cognitive competence can be well explained, both theoretically and empirically, by cultural-religious factors. Protestantism, for example, is seen as education-oriented (compared to other religions and to Catholicism), Confucianism as effort-oriented, in Judaism reading and education were highly valued, etc. (chapter 10.8 [29]). The educational efficacy of Protestantism is supported by numerous studies [25,27].

Table 7
Higher education and science.

	CA mean	Student mean	95 %-level	05 %-level
Quality of universities (05–09)	.32 [.53]	.24 [.51]	.40 [.64]	.06 [.34]
Short duration of study (99–03–16)	.72 [.66]	.72 [.65]	.43 [.28]	.70 [.63]
Professor salary (11–20)	.52 [.62] [.58]	.46 [.65] [.62]	.30 [.50] [.44]	.26 [.40] [.37]
Increase prof. salary (10–20)	.38 [.33] [.51]	.38 [.36] [.53]	.32 [.30] [.48]	.24 [.15] [.23]

Notes. CA mean: cognitive ability mean. Partial correlations in square brackets (west-east and city-rural state differences partialled out); for professors' average salary in the second partial correlation also GDP/c partialled out, always $N = 16$ states.

Table 8
Wealth, society and crime.

	Cognitive ability mean	Student mean	95 %-level	05 %-level
GDP/c (00, 10 17)	−.33 [+.31]	−.42 [+.28]	−.20 [+.30]	−.57 [+.14]
Debts private (04, 05, 06)	−.87 [−.79]	−.83 [−.76]	−.69 [−.71]	−.75 [−.66]
Debts states (00, 10 17)	−.89 [−.76]	−.89 [−.72]	−.73 [−.75]	−.86 [−.58]
Unemployment rate (00–17)	−.34 [−.54]	−.24 [−.50]	−.08 [−.45]	−.11 [−.41]
Welfare receipt (02–16)	−.83 [−.70]	−.80 [−.70]	−.55 [−.60]	−.73 [−.58]
Social gradient	−.74 [−.44]	−.79 [−.39]	−.46 [−.17]	−.87 [−.46]
Crime rate (98–16)	−.77 [−.44]	−.79 [−.45]	−.36 [−.15]	−.81 [−.51]

Notes. GDP/c: gross domestic product per capita 2000, 2010, 2017; Debts private: Household over-indebtedness of private individuals; Debt states: debt of federal states per capita; Social gradient: correlation between socioeconomic status of parents and student attainment; Crime rate crime per capita. Partial correlations in square brackets. $N = 16$ states.

Table 9
Immigration rate.

	CA mean	Student mean	95 %-level	05 %-level	Natives' mean	Migrants' mean
Natives %	.51 [.06]	.60 [.09]	.39 [−.09]	.73 [.22]	−.13 [−.12]	.42 [.31]
Immigrants %	−.51 [−.06]	−.60 [−.09]	−.39 [.09]	−.73 [−.22]	.13 [.12]	−.42 [−.31]

Notes. Data from PISA (2000, 2003, 2006), PIRLS (2001, 2006), and IQB (2009, 11, 12, 15, 16); Natives' mean: mean of students without migration background, Migrants' mean: mean of students with migration background; $N = 16$ German states.

In Germany, however, the pattern is now *reversed* at the state level (Table 10): The proportion of Protestants in a state is (slightly) associated with lower cognitive abilities ($r = -.18$), while more Catholics are associated with higher ones ($r = +.33$). Protestantism seems to have lost its former educational and achievement-promoting effect (pp. 342ff. [29]). The correlation with the proportion of Muslims is negative and considerably higher ($r = -.58$), a pattern that is also found at the international level ($r = -.26$ to $-.71$; Table 10.11 [29]) or in developing countries (such as in Africa [28,71]). Moreover, the correlations of Christian religions are higher with the ability levels of natives (on average $r_{Na} = |.29|$ vs. $r_{Mg} = |.13|$), but for Islam with the ability levels of migrants (on average $r_{Na} = |.06|$ vs. $r_{Mg} = |.42|$). This pattern supports the validity of the findings. Finally, the correlations also help to understand why in Western Europe migrant background has a negative significance for ability – because particularly poorly educated Muslims immigrate here. Causally, the correlations can be explained, among other things, by different educational practices in families.

4.8. Correlations, possible causes and consequences: educational policies and politics

However, all previous correlations are empirically dwarfed by correlations with political variables (Table 11). The percentage of votes for non-left parties in *state elections* is highly positively correlated with cognitive ability (CA) or student achievement (SA) test outcomes ($r_{CA} = +.82$ and $r_{SA} = +.78$), with a non-left orientation of *state governments* being even slightly more highly correlated ($r_{CA} = +.90$ and $r_{SA} = +.88$). The results are robust when removing west-east and city-rural state differences, when considering different levels of ability, and when considering different time periods. Figure A1 in the Appendix shows a scatter plot of votes burgher-civic-conservative-right with the cognitive ability mean. There are two outliers, Bremen and Brandenburg. When both states are removed, the correlation becomes even higher ($r = .89$, $N = 14$). As indicated by the partial correlation (removing west-east and city-rural state differences; $r_p = .88$) the "true" correlation between politics and cognitive ability at the state level is maybe higher. However, the correlations for political orientation of state governments declined in the last decade because a tendency towards burgher-conservative-right election results can now lead to progressive-left state governments (in Baden-Württemberg, Saxony, and Thuringia; the other parties avoid any cooperation with the right-wing AfD).

When looking at the individual parties, the CDU/CSU-SPD contrast is particularly important, empirically because high and robust correlations are found here, and substantively because these parties determined the prime ministers and the guidelines for education policy for decades. If taken prime ministers as a political indicator, the correlation with the cognitive ability mean is $r = .86$ ($r_p = .83$; not in the table). The correlations do not depend on chosen indicators. Tendentially, politics seems to be more important for weaker students, correlations with the 05 % threshold are always higher than with the 95 % threshold. If all correlations are used as absolute values the means are using Fisher's z-transformation $r_{M95\ \%} = .51$ and $r_{M05\ \%} = .65$. Strong students are able to realize their potential

Table 10
Religion and culture.

	CA mean	Student mean	95 %-level	05 %-level	Natives' mean	Migrants' mean
Protestants %	−.18 [−.28]	−.26 [−.33]	−.48 [−.45]	−.23 [−.17]	−.02 [−.31]	.06 [−.04]
Catholics %	.33 [.27]	.29 [.34]	.08 [.24]	.20 [.32]	.51 [.33]	.36 [.05]
Muslims %	−.58 [−.25]	−.66 [−.27]	−.44 [−.11]	−.77 [−.36]	.02 [−.09]	−.49 [−.34]

Notes. Partial correlations in square brackets: West-east and city-rural state differences partialled out.

Table 11
Politics (elections and governments at the state level).

Parties and governments	Cognitive ability mean	Student mean	95 %-level	05 %-level
Votes and governments in left-right scale				
Votes burgher-civic-conservative-right	.82 [.88]	.78 [.90]	.55 [.57]	.66 [.84]
CDU/CSU minus SPD contrast	.89 [.90]	.89 [.91]	.69 [.62]	.83 [.88]
Governments burgher-civic-conservative-right	.90 [.91]	.88 [.93]	.64 [.63]	.80 [.91]
Votes and governments in left-right scale 2000–19				
Votes burgher-conservative-r. 2000–2019	.87 [.79]	.86 [.81]	.62 [.54]	.78 [.70]
CDU/CSU minus SPD contrast 2000–2019	.83 [.77]	.83 [.79]	.62 [.55]	.78 [.76]
Governments burgher-conservative-r. 2000–2019	.64 [.49]	.63 [.50]	.40 [.31]	.59 [.48]
Party results				
Left (Linke, PDS) 2000–2019	.02 [.02]	.11 [.04]	.23 [.03]	.22 [.11]
Greens (Grüne) 2000–2019	−.29 [.14]	−.37 [.12]	−.17 [.22]	−.48 [.06]
SPD 2000–2019	−.66 [−.70]	−.69 [−.71]	−.68 [−.60]	−.68 [−.68]
CDU/CSU 2000–2019	.73 [.71]	.69 [.74]	.31 [.33]	.61 [.71]
AfD 2000–2019 (since 2013)	.21 [.11]	.31 [.16]	.45 [.34]	.31 [−.10]

Notes. votes burgher-civic-conservative-right: votes for burgher-civic-conservative-right parties (since 1959/1990 or from 2000 on), specifically 100 minus votes for SPD, Greens, Left, and Pirates; Governments burgher-civic-conservative-right: state government burgher-civic-conservative-right (since 1959/1990 or from 2000 on); party results: Order top-bottom from political left to right; FDP as center party not displayed; $N = 16$ states.

more independently of the respective educational policy. The same high-ability-independence pattern was found for educational environment (Table 6).

An exception of the left-right pattern are the tiny positive correlations with votes for the party Die Linke (The Left; $r = .02$). All correlations including the partial correlations are small positive. For the Greens, the negative correlations turn into positive ones when west-east and city-rural backgrounds are partialled out. In the past, both parties were far less important for education policy than CDU/CSU and SPD. And both parties have strong geographic tilts: The Greens are strong in cities, The Left in the east. Such regional patterns may bias correlations. All progressive-left groups combined result in stable negative correlations.

The correlation between politics and cognitive ability is very high ($r > .80$; Table 11). A possible cause can be found in the high correlation between politics, educational policies and student achievement resp. cognitive ability (Table 12). Burgher-civic-conservative-right state governments (GovR) and electoral outcomes/votes (ElR) are associated with burgher-civic education policies ($r_{GovR} = .87$ and $r_{ElR} = .75$, row 1 in Table 12), e.g., more likely to favor centralized testing ($r_{GovR} = .78$ and $r_{ElR} = .68$), tracking at a young age ($r_{GovR} = .40$ and $r_{ElR} = .50$) and more traditional instruction ($r_{GovR} = .52$ and $r_{ElR} = .31$). Figure A2 in the Appendix shows a scatter plot of governments burgher-conservative-right with the conservative educational policies mean. There are three outliers, Bavaria, Berlin and Hamburg. When they are removed, the correlation is still $r = .86$ ($N = 13$; vs. $r = .87$, $N = 16$). Thus, outliers do not bias the positive correlation between governments and corresponding education policies.

However, not all correlations are high and stable, e.g., grading and homework in elementary school. The more global variables, e.g., Klafki proximity and achievement structure, the more stable are the correlations. Progressive-left state governments and electoral outcomes are associated with higher comprehensive school rates ("Gesamtschule", $r_{GovR} = |.76|$ and $r_{ElR} = |.77|$). These results show that politics controls education policy – not surprisingly. However, that is not all. For example, looking at the positive correlation between burgher-civic-conservative-right electoral outcomes and student discipline ($r = .26$ and $r = .45$), and considering that the political orientation of a state government depends on the political orientation of its state population, there is probably also a direct

Table 12
Politics and burgher-civic-conservative-right educational policies.

	Governments burgher-conservative-r	Votes burgher-conservative-r	Governments burgher-c-r 2000–2019	Votes burgher-c-r 2000–2019
Conservative educational policies mean	.87 [.91]	.75 [.85]	.65 [.49]	.82 [.74]
Central exams (1995)	.78 [.84]	.68 [.81]	.54 [.50]	.68 [.72]
Tracking at a young age	.40 [.26]	.50 [.37]	.48 [.38]	.51 [.40]
Discipline	.46 [.28]	.26 [.14]	.34 [.01]	.45 [.08]
Comprehensive school rate (i)	−.76 [−.66]	−.77 [−.75]	−.46 [−.24]	−.76 [−.61]
Gymnasium rate (i)	−.58 [−.45]	−.60 [−.41]	−.27 [.10]	−.54 [−.29]
High-school rate (Abitur) (i)	−.81 [−.72]	−.64 [−.47]	−.45 [−.11]	−.63 [−.27]
Achievement requirements	.80 [.81]	.67 [.70]	.41 [.26]	.71 [.62]
Traditional instruction (PISA)	.52 [.62]	.31 [.48]	.48 [.51]	.45 [.46]
Grading from grade 2 onwards	.11 [−.13]	.02 [−.10]	.35 [.21]	.24 [−.01]
Homework in grade 4	.09 [−.13]	.04 [−.09]	.31 [.20]	.25 [−.01]
Klafki proximity (i)	−.81 [−.86]	−.85 [−.81]	−.45 [−.22]	−.77 [−.63]
Achievement structure	.79 [.74]	.62 [.60]	.36 [.14]	.57 [.39]

Notes. Conservative educational policies mean from Table 6. Governments burgher-conservative-r: left-right government classification 1959–2019; votes burgher-conservative-r: votes for burgher-civic-conservative-right parties 1959–2019; same for 2000–2019. (i): scales inverted for the dimension educational policies mean; $N = 16$ states.

effect via a non-political path, e.g., via education in families on students' achievement. Student and parent attitudes are reflected in children's behavior at school and in election results.

4.9. Path models

The most important factors turned out to be the political orientation of state governments and education policy (Tables 11 and 12). In addition, there was the proportion of students without (or with) a migrant background (Table 9). It was described and empirically proven by correlations that condition variables are interrelated (the pattern of burgher-right vs. progressive-left educational policies with election results; Table 12).

In a regression analysis that takes this relationship into account to predict the criterion variable student achievement studies mean, educational policy, the political orientation of state governments and the proportion of students without a migration background show a positive effect (Fig. 3). The effects of policy ($\beta = .41$) and education policy ($\beta = .45$) are roughly equal, and "native background" has a smaller effect ($\beta = .21$) on state differences. Thus, ability differences across states seem to be primarily politically determined, with education policy being dependent on the voting decisions of the respective state population. Interactions and indirect effects among them cannot be analyzed by a simple regression. This dependency is displayed in the next model (Fig. 4).

Dependencies between state government and education policy are shown in the path analysis in Fig. 4. The analysis was done with Mplus; the fit was very good: Comparative Fit Index *CFI* = 1, Standardized Root Mean Residual *SRMR* = .04. Data as of September 2019. The political left-right-orientation of a state government depends on the election outcome ($\beta = .93$). Left (as left categorized) electoral results lead to left (as left categorized) governments, right electoral results to right governments. That is not at all surprising. The high path coefficient underscores the validity of the data and the analysis. The political orientation of the state government influences education policy (burgher-civic-conservative-right educational policy $\beta = .87$, education quantity $\beta = .38$). Governments that are more conservative opt for more traditional education policies, more progressive governments opt for more progressive ones. That, too, is not surprising and underlines the validity of the data. The impact on right-left categorizable education policies is much stronger than on the quantity of education, for which we previously had no theoretical background (i.e. that conservative politics favor quantity of education). Both in turn have positive effects on the average result in student achievement studies ($\beta = .63$, $\beta = .29$) standing for higher cognitive competence of students. The education policy effect ($\beta = .63$) is considerably larger than the education quantity effect ($\beta = .29$). It should be kept in mind, however, that we are talking about state differences, not about the development of individuals. For individual development, the quantity of education may be more important.

Theoretically open is the remaining direct effect of electoral outcomes on student achievement ($\beta = .23$; Fig. 4). In structurally equal path analyses but with other variables (e.g., homework instead of quantity of education) the direct effect of electoral results on student ability was even larger. Independent of educational policies, there seem to be direct effects, these could work via student diligence, school-positive behavior (discipline), reading to the child, cultural interests, completeness of the family, educational values and educational level of parents, etc. The positive effect of burgher-civic state government on educational quantity ($\beta = .38$, $r = .38$) was initially not expected. Educational quantity includes the four variables of adult educational level, kindergarten attendance, amount of instruction, and no teacher shortage (see Table 6). Presumably, an emphasis on achievement leads to increases in these aspects as well.

So far, the positive "effect" of certain policy orientations on (youth) ability may also be caused by a reverse effect of citizens' ability on political opinions and decisions. This can be tested in a cross-lagged analysis comparing the effects of politics on ability with those of ability on politics. In a final analysis, interaction processes between policy and cognitive ability are examined longitudinally over 5 decades (Fig. 5). The figure reports the mean of longitudinal analyses of cross-lagged effects from the mid-1960s to 2019; variations in

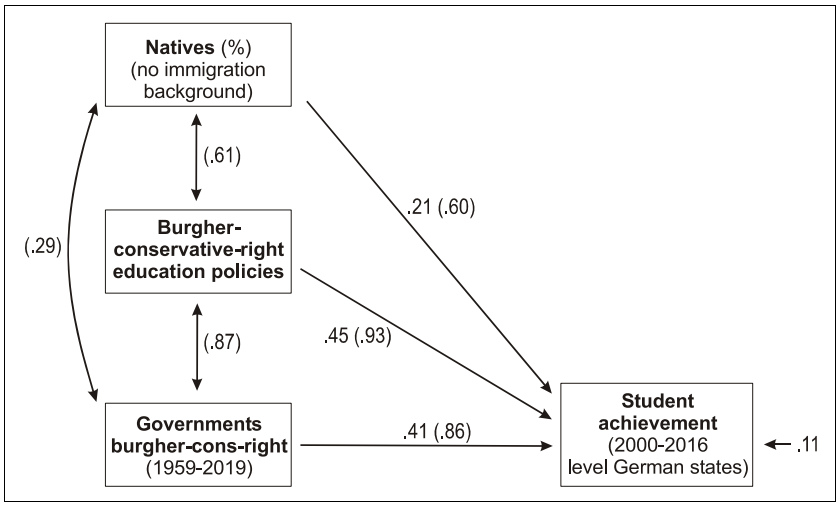


Fig. 3. Regression analysis (standardized path coefficients and in parentheses correlations, $N = 16$ states).

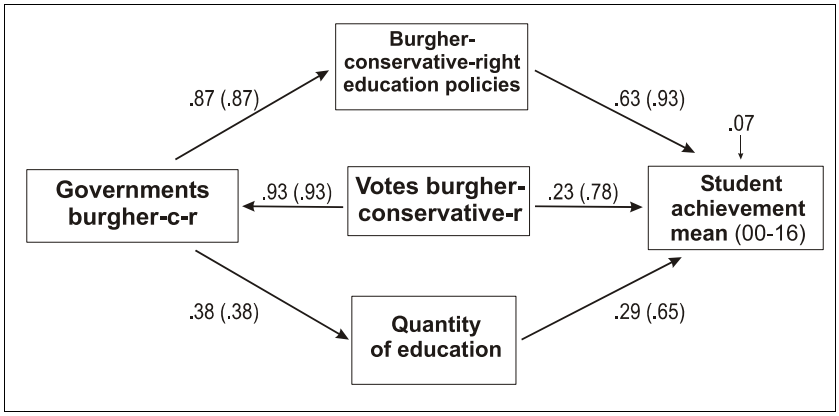


Fig. 4. Political milieu (votes left-right in German state elections; in the center), state politics (governments left-right; on the left), educational policies (left-right; at top), quantity of education (at the bottom) and student achievement (average; on the right) ($N = 16$ German federal states).

results that may occur due to small state samples are averaged out in this way ($N = 16$ resp. 11 states). The result shown is based on the weighted average of eight analyses over different time periods, from 1969 to 2000 with general or Gymnasium student sample, 1971 to 2006, 1973 to 2006, 1991 to 2006, 1992 to 1998, 2000 to 2016, and on different student achievement tests.

A burgher-civic-conservative-right orientation of state government promotes student achievement over time, which stands for cognitive competences, for thinking and knowledge ($\beta = .45$). Student achievement as a measure of cognitive competences, in turn, has a (smaller) effect via political orientations and elections on state government orientations ($\beta = .32$): In Germany, a bourgeois culture seems to make people smart and smartness makes voters bourgeois.

5. Discussion

The aim of the study was to describe differences between German federal states in student achievement and cognitive ability and to determine their possible causes and consequences. The various student achievement and competence studies come to a largely uniform ability pattern for the 16 German federal states. The homogeneity of the results is high (Cronbach-$\alpha = .93$). Countries with high-performing students in reading also do well in math; if they do well in fourth grade, they also do well as 15-year-olds; if they do well on the IQB, they also perform well on PISA. Students in the south perform better in cognitive ability measures, an unusual result compared to international patterns. A small peculiarity concerns the foreign languages (the students in the east achieve weaker results) and the natural sciences (the east is stronger), which is probably still a (fading) reminiscence of former differences in school education between west FRG (BRD) and east GDR (DDR). The size of the differences between the states is substantial, reaching half a standard deviation and about 14 months of learning time in school, respectively, with Bremen and Bavaria at the lower and upper ends.

The state differences correlate with educational characteristics. Overall, a "stricter", more conservative and achievement-based education policy is associated with better outcomes, behind which lie more burgher-civic-minded state governments and electoral decisions by state populations. The pattern does not depend on the operationalization chosen, i.e., whether one looks at individual parties (e.g., votes for CDU/CSU and SPD), whether one combines them into the two blocs left vs. right, whether one contrasts CDU/CSU and SPD, whether one puts state governments into a left-right scheme, or whether one takes prime ministers.

A lower percentage of immigrant students also helps explain the competence gap across states, but education policy is more

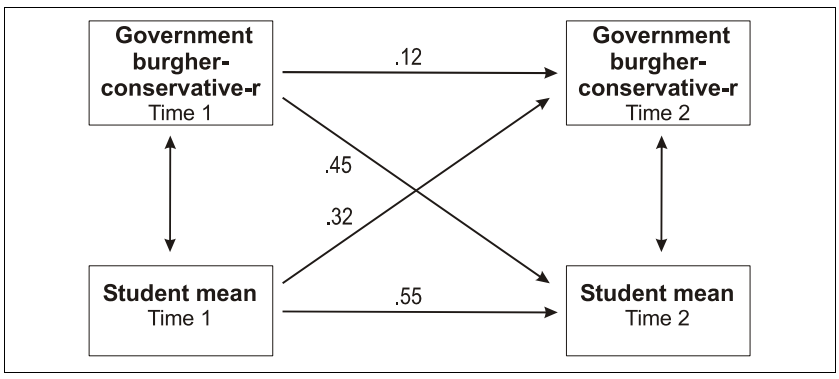


Fig. 5. Reciprocal effects between politics and abilities ($N = 16$ German states, average of eight single longitudinal analyses, averaged standardized path coefficients).

important. Differences in ability, in turn, have social consequences, such as in the handling of money or the law, but differences in ability are not the only causal factor for state differences in such variables. The higher the ability level, especially of low ability groups (05 %-level), the fewer people suffer from debt, the less debt a state has, the less welfare has to be paid, the less the ability level of children depends on the SES of their parents, and the fewer people suffer from crime. On the other hand, the higher the ability level of intellectual classes (95 %-level, also called smart fraction, gifted, high ability students), the better the universities and the higher the average salaries of professors in a state (GDP/c controlled). This underlines the validity of the measures in general, of the pattern of results (95 %-level vs. 05 %-level) and the specificity of the effects as a function of the selected social groups (for international comparisons: pp. 123ff. [29]). A similar supportive result was found for percentage of migrants in a state: percentage of migrants is more relevant to migrant test scores (and for the low ability groups). Finally, the proportion of Muslims is more important for the ability level of migrants and for lower-performing groups. Although we only have 16 states and although there is some doubt about the percentages on migrant background in particular, and although there may be some variation between states in the origin of migrants, the numbers seem to be meaningful (but the migration effect might be somewhat underestimated, see Limitations).

Robust correlations, on the one hand different achievement and ability measures, upper and lower thresholds (95 % and 05 %), native and migrant values, on the other hand different indicators of educational policy, of policy choices and conditions, of societal and economic characteristics, speak for stable patterns. Longitudinal analyses point to interaction effects, i.e., educational policy influences ability, but ability levels in turn influence policy orientations and decisions (Fig. 5). Testing the robustness of effects across different variables and time periods is a strength of the study (see also the correlation at the level of 401 electoral districts; chapter 7).

The result is scientifically interesting in itself, both because it replicates proven findings from other studies with different levels of data, e.g., relevance of kindergarten attendance, no teacher shortage, central exams, discipline, and because it reveals rather uncomfortable truths for many: a more stern, traditional, and achievement-oriented education policy leads to better results. Educational policy decisions, if better student achievement and higher cognitive competence are desired, could take this into account. Intended effects and side effects should be considered and balanced.

6. Limitations

Methodology: As always on the macro-social level, causal interpretations cannot be safeguarded against every objection. For example, intervening processes such as specific student learning behaviors or family education would need to be supplemented. Additional determinants such as cultural-religious orientations and wealth play a role. There is no theory for genetic factors in Germany, and our study shows that known genetic markers of cognitive ability have no influence, but genetic effects cannot be ruled out, e.g. ability-related migration between regions [2,35,72]. Finally, we are dealing with a small sample of 16 states, which may lead to multicollinearity and unstable effects. A relatively small sample leads to greater variation in results depending on the variables chosen, the additional variables in a model, and the complexity of the models. We attempted to address this problem by pooling results from different studies, variables, and time periods. We checked scatter plots, correlations are similar when outliers are excluded. While single variables sometimes show unstable results (e.g. homework in primary school), the overall scales (dimensions based on multiple variables) show robust results. Due to the small sample of states and the analyzes only within one country, results that contradict findings at other levels or in other countries (e.g. the negative correlation between high-school rate and student ability; Table 6) should not be generalized.

Immigration: Due to the groups of students who withhold information on migration background but have low test scores, the achievement level of migrant students is likely to be overestimated and the difference between natives and migrants underestimated. This means that the migration effect is probably somewhat stronger than observed (e.g., instead of $\beta = .21$ could be a "true" $\beta = .27$; Fig. 3). Nevertheless, the migration effect is unlikely to reach the scale of the education policy effect. However, the migration effect is already not insignificant: While in 1992 the younger generation (male conscripts) in East Germany lagged significantly behind in cognitive ability (determined by the army tests: west IQ 100.29, east IQ 94.55), the pattern reversed in the student assessment studies from 2000 to 2016: west IQ 98.74, east IQ 100.07. A lag of about 6 IQ points resulted in an advantage of about 1.5 IQ points. Eastern Germany's lead is due to the lower proportion of immigrants (around 30 % vs. 7 %): Looking only at the ability level of natives, the west is still ahead, but with only a small lead of 1 IQ point compared to the 1992 6 IQ points gap (west IQ 101.87, east IQ 100.88; results of student achievement tests displayed in an IQ scale).

Generalizability of the burgher-progressive or right-left pattern: For Germany, it can be clearly stated that more burgher or right-wing conservative policies are associated with higher cognitive achievement levels of students in school. We tried to show that this has to do with education policies that are more strict, and more focused on achievement (rather than equality and the like). Conversely, there was empirical evidence that cognitively competent populations tend more toward burgher-conservative-right-wing politics, i.e., are more likely to vote CDU/CSU. The descriptive findings are robust; one can discuss the causes. However, one must warn against transferring this to other countries and epochs without further consideration. Only those political conditions are associated with higher abilities that create environments more conducive to cognitive development. And this can vary depending on the historical constellation. Thus, the communist-centralist ("left") policies of the Soviet Union in Central Asia (Uzbekistan, Kyrgyzstan, etc.) were certainly more cognitively beneficial than traditional Islamic-influenced ("right") policies there (e.g., Luria, 1976/1974 [73]). Anywhere, for example, where education for girls is hindered, or where only traditional rote learning is practiced, test scores are likely to be worse, just as when there is no longer sufficient attention to discipline, central exams, strenuous learning, or effective direct instruction.

Comparison with other researchers: It would be good to have an independent analysis by other researchers using different methods, as is the case at the international level by educational economists (e.g., Eric Hanushek and Ludger Woessmann). However, it does not seem very likely that this research will take place.

7. Few or no research – what can be the reasons?

> "A great deal of attention has been given to epistemology (the study of how we know) when 'how or why we don't know' is often just as important, usually far more scandalous, and remarkably undertheorized." (Proctor & Schiebinger, 2008, p. vii [74])

According to my considerations, there are four reasons why there is no research on the differences within Germany:

(1) First, in psychology as a science of human beings, there is the ideal of the small, clean, well-controlled study with a large sample of individual data. The societal level, which is undoubtedly difficult to research with all its imponderables, associated covariates, and uncontrollable background factors is rather left to sociology and economics. Furthermore, a sample with only 16 observations is usually not considered analyzable. For example, one rule of thumb is that for regression or path analyses, the sample size should be about 10–20 times the number of parameters, i.e., even for two to three variables, the sample size would be too small. Others recommend 100 to 200 observations [75]. But what alternatives are there for analyzing differences among 16 states? Alternatively, how can one explain the observably high and robust correlations with political conditions and education policy?

One reasonable option is that other authors attempt to replicate the research using similar variables and models. Another option is to perform bootstrapping. Finally, for certain research questions, 401 electoral districts can be selected instead of 16 states. We mapped their election results to the intelligence test scores of 1998 male conscripts in 83 district recruiting offices (see Fig. 2). Again, the more votes for the burgher-conservative-right parties, the higher the intelligence ($r = .32$, single parties: CDU/CSU $r = .48$, SPD $r = -.40$). Note that not only has the sample size increased (83 districts instead of 16 states), but the unit of analysis has also changed, which is a kind of cross-validation of the policy ability pattern mentioned above.

(2) Another reason could be the political dependence of school-related research. Any study involving pupils in schools must be approved by the authorities. Germany no longer conducts the state-based PISA studies. It is hard to imagine that the state-based IQB studies could continue to be conducted by these researchers if they found relationships of the results to the decisions of the political parties in power in Germany. The researchers would jeopardize the conduct of upcoming studies, the existence of their institutes and the professional future of a large number of employees. In such a situation, researchers cannot say that the emperor is naked.
(3) In addition, psychology is a field of science dominated by researchers with certain beliefs in the political realm, specifically "progressive" and "left-wing" beliefs (e.g., for the U.S., left to right is 14:1 [76]). There is low political diversity in psychology. This may lead to a variety of biases, such as in research questions that are asked, in papers that are submitted and published, in outcomes of peer review, in the allocation of external funding, in ethics committee decisions, and in appointments [77–81].
(4) A similarly skewed political mood also prevails in the German media [82,83] and in the most used encyclopedia, Wikipedia, which influence science [84–86]. All factors combined, i.e., incentives and pressures from the authorities, from fellow scientists and their prevailing beliefs, and from the media, as well as scientists' own beliefs, can prevent research in this area from being carried out, at least when there is a risk that it will not produce the expected results [87]. Intentionally, non-knowledge is generated [74].

8. Conclusions

The study sought to answer the question of why the German states get such considerably different results in student achievement and cognitive competence studies. Largely independent of the paradigm of the study (student achievement or intelligence), individual study (PISA, PIRLS or IQB), age of the tested (from 9/10 to 18 years), chosen dimension (reading, mathematics or science) or year of the survey, the southern states are ahead and the northern and city states are behind. The differences reach 51 student assessment scale or 8 IQ points equivalent to 14 months of learning gains in schools. Three factors turned out to be decisive: a more traditional and achievement-based education policy, e.g. central exams and high requirements, behind this, rather burgher-conservative state policy and society, and finally the percentage of students with no migration background.

Data availability statement

The data will be uploaded to ResearchGate with the publication (www.researchgate.net/profile/Heiner-Rindermann).

Funding

The publication of this article was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 491193532 and the Chemnitz University of Technology.

CRediT authorship contribution statement

Heiner Rindermann: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal

analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Heiner Rindermann reports financial support was provided by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

A1 Method

A1.1 Student achievement and cognitive ability studies

Student achievement depends on teaching and school characteristics [64]. But it also depends on family and neighborhood characteristics and especially on genes as well [88,89]. Achievement is important as an indicator for ability of a person, of a student subgroup, e.g., with or without immigrant background, for quality of an institution, e.g., school or school system, and for average ability of a nation, e.g., US students or the British population.

Student achievement tests are a good indicator of cognitive ability, additionally, causes and consequences of student achievement and intelligence are similar [90,91]. Similar to intelligence tests, PISA scales show a strong *g* factor (about 70 % of variance [92]). PISA scales are also highly correlated with intelligence test scales (the German CogAT, at the observed level $r_o = .47$, at the latent level $r_l = .86$ [93]). A mean value calculation highlights the *g*-character, specific factors are averaged out. Competence levels are variable and can be modified by personality and environment. Since we are working here with data from student achievement studies, which in the case of the IQB studies were also developed with a closer orientation to the curriculum, greater malleability depending on educational policy is to be expected. We drew on the following sources:

PISA (Programme for International Student Assessment) from 2000, 2003 and 2006, 15-year-old students in reading, mathematics and science, 2003 also problem solving; dimensions (tasks) are not closely related to the content taught in school, but represent a more general "literacy". For the years 2009 to the present, results have not been published for the 16 German states. All student achievement tests use scales with a mean of 500 and a standard deviation of 100. Sources were German PISA reports (e.g., from the PISA-Konsortium in 2008 [94]).

PIRLS (Progress in International Reading Literacy Study, abbreviated in German as "IGLU") from 2001 to 2006 (fourth graders in reading, like PISA not close to school curricula, "literacy" [95]). No results have been available for the German states since 2011.

IQB studies (Institute for Educational Quality Improvement, in German „Institut zur Qualitätsentwicklung im Bildungswesen") from 2009, 2011, 2012, 2015, and 2016 [96]. Dimensions vary depending on content and grade level (fourth- and ninth-graders). Pupils were tested in reading-German, orthography, foreign languages, mathematics or science. The dimensions are curriculum related and close to school. Foreign languages were not considered in the averaging process in our study, reasons: first foreign language varies depending on the federal state and school, i.e. English, French or Latin; foreign languages are not otherwise included in studies of student achievement.

No German state-level values are available from *TIMS* studies (Trends in International Mathematics and Science Study). Further regional and single studies within Germany (such as *VERA* or *Markus*) were not included; reasons: not carried out in all states, not frequently repeated.

German armed forces (Bundeswehr) *intelligence test* 1998, only young men who did not do conscientious objection, the test is part of the psychological examination (the results shown in a diagram were converted to the IQ scale [35]).

The following results, presented in tables, were taken from the reports (with the exception of the Bundeswehr data): the state means in the single dimensions and the results for the low ability as well as high ability (intellectual class) students (05 % and 95 % percentile thresholds); means for high school students, in German "Gymnasium"; and means for students without and with an immigrant background, all nations and different degrees of own and parental immigrant background were combined here, further called "natives" and "migrants".

There were often problems with the two last values. In the IQB studies, for example, the results for natives and migrants have changed somewhat (e.g., differences for the year 2009 in the 2010 and 2015 published reports). Another problem concerns a group that is usually not mentioned in the studies, namely pupils who do not provide information on their migration background. This problem can also be found in studies at the international level. Those pupils usually score weakly in the achievement scales. On average, 12 % of students in Germany did not provide information about their migration background; in Hamburg, the number was as high as 24 %. Since the proportion of migrants and the percentage of the unknown group are positively correlated ($r = .66$), there is no apparent motivation why (weak) natives should be more likely than (weak) migrants not to report their origin, and migrants have weaker results on average, the unknown weak group presumably includes an above-average number of students with a migrant background. On average, this leads to an overestimation of the values by .42 IQ points or 2.80 SASQ points or to an underestimation of the native-migrant difference (is: 8.08 IQ or 53.87 SASQ, after a possible correction: 10.62 IQ or 70.80 SASQ). We did not correct for this (unlike internationally, where the discrepancies between reported and calculated values are even larger [69]). However, it should be noted that the given data probably *overestimate migrant scores* by about 2 IQ or 15 SASQ.

By averaging across different dimensions within studies (not for PIRLS, since only one dimension was measured), an overall mean was calculated for each study and survey year. In the next step, for each study (PISA, PIRLS, IQB), a study mean was calculated over different years after common standardization. Finally, the student assessment studies (weighted five times because far more studies from more years and more recent) were combined with the single Bundeswehr study.

A1.2 Older ability studies for comparison

The following data from older studies were used to test the stability of the federal state patterns over time:

- *Reading ability* (comprehension, literacy) in grammar school (high school, "Gymnasium") in 1969 documented by Baumert et al. (2003, p. 321) [52].
- *Student achievement in arithmetic* and *German* in 1971, measured via the RT2 (Rechentest für 2. Klassen, Arithmetic Test for 2nd Grades) and the AST2 (Allgemeiner Schulleistungstest für 2. Klassen, General Student achievement Test for 2nd Grades, numeracy and text tasks, spelling, and vocabulary; Flor et al., 1992, pp. 18, 20 [53]). The age of the pupils was between 7 and 8 years.
- *German* and *mathematics* abilities of high school graduates ("Abiturienten", age approx. 19 years) in 1973 (Trost, Pauels & Schneider, 1976) [54].
- *General student achievement* in 1991 in the Allgemeine Schulleistungstest für 4. Klassen (General Student achievement Test for 4th graders; Flor et al., 1992, pp. 13, 29 [53]). Age of pupils between 9 and 10 years.
- *German armed forces* (Bundeswehr) *intelligence test* 1992 (Ebenrett et al., 2002, p. 3) [35].
- *Preliminary medical examination* (Physikum) scores of university medical students in an objective and uniform examination in medicine from 1994 to 2000 (Robra & Schmitt, 2001 [55]). Age around 20–24 years.

If multiple scores were available, they were averaged, if necessary, after standardization. In order to compensate for possible methodological problems of the old studies, a *g* factor was formed from them (the first unrotated factor, missing values replaced via mean substitution).

A1.3 Educational factors

The following educational, instructional, student, school system, and educational policy characteristics were used as potential explanatory variables for state differences in ability:

- *Educational level of adults* from IQB (parents 2011 and 2016) and of adults in general (e.g., Stanat et al., 2017 [96]; for 2017, Destatis, www.destatis.de). The higher the value, the longer on average adults have attended educational institutions.
- *Kindergarten attendance* (3−5 years) and *crèche attendance rates* (0−2 years) 2012–2017 (Destatis, 2018b [97]). The higher the value, the more and longer children have attended kindergarten.
- *Amount of instruction* according to PISA 2000, PIRLS 2001, and 1991 (teaching hours per year; e.g., Baumert et al., 2003 [52]; Spiegel, 1991, pp. 54−55 [98]).
- *No teacher shortage*: a sufficient number of teachers and an appropriate deployment of teachers according to their training for the subject in question (PISA 2000, 2006, IQB 2011; e.g., PISA-Konsortium, 2008 [94]).
- *Central examinations* in 1995 in schools (Hauptschule, Realschule, Gymnasium; Büchel et al., 2003 [62]). This means that all students take a uniform exam on the same day, which is evaluated externally. This variable has two meanings: on the one hand it stands for the narrower (older) characteristic of the central examinations, on the other hand it stands for a broader characteristic of achievement orientation.
- *Tracking at a young age* (streaming in different schools) with young-as-high numbers (retrieved from state ministries of education). Tracking is tied to past achievement at school, but not necessarily very closely to it (e.g. not linked to results in institutionally prescribed objective tests).
- *Student discipline*: not truanting or being late ("15-year-olds who never missed class in the last two weeks of school", PISA 2000; Baumert et al., 2003 [52]), data are given for $N = 14$ states.
- *Comprehensive school rate*: percentage of students in comprehensive schools ("Gesamtschule"; PISA 2000, 2006 and IQB 2009; e.g., PISA-Konsortium, 2008 [94]). Comprehensive schools do not select pupils on the basis of achievement.
- *Gymnasium rate*: percentage of students in grammar schools from PISA and IQB (e.g., PISA-Konsortium, 2008; Stanat et al., 2017 [94,96]). Gymansiums lead to the baccalaureate (Abitur) and then to university studies.
- *High school graduation rate* (1995–2000; "Abiturquote") from Diefenbach and Klein (2002) [99].
- *Achievement requirements* (standards) are good and (reasonably) high: Percentage of parents who do not consider the achievement requirements of school to be too low or much too low (PISA 2000; Baumert et al., 2003 [52]), $N = 14$ states.
- *Traditional instruction* from PISA 2006, e.g., few experiments done by students themselves (PISA-Konsortium, 2008, p. 280 [94]), standing for more teacher-directed instruction.
- *Grading* in the second grade of elementary school (Tricarico, 2015 [100]), i.e. pupils receive grades at a young age (and not just in grades 3 or 4).
- Amount of *homework* in reading in grade 4 (PIRLS 2006; Bos et al., 2008 [95]).
- "*Klafki proximity*" (Böhm, 2008, p. 93 [101]): Wolfgang Klafki (1927–2016) was a renowned German education researcher and progressive educational reformer. Christian Böhm (2008) [101] rated the educational policy of the 16 German states regarding

their affinity to the progressive educational reform ideas of Wolfgang Klafki. Reversed this Klafki proximity rating (general progressive education policy) stands for burgher-conservative orientation in educational policies of a German state.
- *Achievement structure* by educational policy (means of structure specification, content specification, and achievement control specification in 1997, Cronbach-$\alpha = .76$; according to Below, 2002 [102]). Structure specification stands for Gymnasium attendance in younger age, content specification for foreign language compulsory until high school graduation (Abitur), achievement control specification for central examinations etc. The variable represents high achievement orientation and control by the state.
- *"Burgher"* (or civic, conservative, bourgeois, right) educational policy (see also Fend, "*The pedagogy of neoconservatism*", 1984; Weinert, 2001; Below, 2002 [102–104]) has been summarized based on: central examinations, tracking at a young age, higher student discipline, comprehensive school rate (reversed), grammar school rate ("Gymnasium", reversed), high school graduation rate ("Abitur", reversed, 1995–2000; Diefenbach & Klein, 2002 [99]), high achievement requirements, traditional instruction, grading from the 2nd grade (at a younger age compared to starting in the 3rd or 4th grade), a lot of homework in reading in grade 4, Klafki proximity (reversed), achievement structure (Below, 2002 [102]). The global educational policy scale has a reliability of Cronbach-$\alpha = .84$.
- *Educational quantity* covers adult educational level, kindergarten attendance, amount of instruction, and no teacher shortage (Cronbach-$\alpha = .55$).
- *Migration background* (percentage of students): data from PISA (2000, 2003, 2006; e.g., PISA-Konsortium, 2008 [94]), PIRLS (2001, 2006; e.g., Bos et al., 2008 [95]), and IQB (2009, 11, 12, 15, 16; e.g., Stanat et al., 2017 [96]). Between studies definitions differ slightly and over time there is an increase, therefore averaged after standardization (Cronbach-$\alpha = .98$).

Except for discipline and achievement requirements, data are always available for $N = 16$ states.

A1.4 University and research

University, student, and research characteristics can be causes and consequences of abilities. The following variables were considered for all 16 German states:

- *Quality of universities* based on data from CHE 2005 and 2009 on research quantity and quality (Centrum für Hochschulentwicklung/CHE, 2009 [105]) and the Excellence Initiative 2006 (graduate school funding, clusters of excellence, university of excellence), Excellence Initiative data relative to population size (www.dfg.de); Cronbach-$\alpha = .72$.
- *Short duration of university study* reversed in 1999, 2003, and 2016, i.e., high numbers represent students achieving in short time a university degree (Wissenschaftsrat, 2005; Destatis, 2018a [106,107]). The older (1999, 2003) and younger (2016) indicators of duration of study do not correlate.
- *Professors' average salary*: Salaries of W grade (W2, associate professor, and W3, full professor, averaged) in 2011 (base salary), 2015 (average salary), 2017 (average salary), 2019 (base salary), and 2020 (base salary), Cronbach-$\alpha = .81$. *Salary increase* is 2020 minus 2010 in base salary (source: German university professors' journal *Forschung & Lehre*, 11/8, 17/2, 19/8, 20/6).

A1.5 Economy and society

Internationally, there are interactions between cognitive human capital (cognitive ability) and the economy and society with, on average, larger effects of ability on the economy and society than vice versa (e.g., Rindermann, 2018 [29]). The following characteristics were considered for the 16 German states:

- *Gross domestic product per capita* (GDP/c) 2000, 2010, 2017, source is Institut der deutschen Wirtschaft (2019) [108].
- *Private debt*: Debt overload of private households from the years 2004–2010, 2014 and 2018, in percentages (Creditreform, 2018; Schufa, 2010 [109,110]).
- *Public debt* of the federal states per capita 2000, 2010, 2017 (Destatis, 2018c [111]).
- *Unemployment rate* 2000, 2002, 2010, 2015, 2017 (Destatis, 2018c [111]).
- *Welfare receipt* (Hartz IV benefits) per capita 2002, 2012, 2013, 2016 (Bertelsmann, 2015; Destatis, 2018c [111,112]).
- *Social gradient*: Correlation between socioeconomic status of parents and student achievement test result (PISA 2006, IQB 2009, 2011, 2012, 2015, 2016; e.g., PISA-Konsortium, 2008, Stanat et al., 2017 [94,96]). A high correlation indicates more dependency of student achievement test results on parental SES.
- *Crime rate*: Frequency number for total crime per capita 1998, 2000 2002, 2014, 2015, 2016 (Bundeskriminalamt, 2017 [113]).

Reliabilities are usually very high, e.g. for GDP/c Cronbach-$\alpha = 1$, for crime $\alpha = .96$.

A1.6 Religion and culture

Internationally and based on correlations and path coefficients, culture (religion) proved to be an important determinant of educational differences [29]. For Germany, the percentages for Protestants, Catholics and Muslims for 2011 were taken from Statista from 2018 [114].

A1.7 Politics: State elections and state governments

In Germany, school policy is a state matter. State governments influence the educational system, schools, teaching and students through educational policy decisions. State governments are elected by parliaments, which in turn are elected by the citizens of a state.

Characteristics of citizens, their orientations and abilities, determine government policy over time, and government policy influences abilities through educational policy decisions over time. The following political characteristics were considered:

- *Election results in state elections*: 2000 to 2019, for the parties Die Linke (PDS, The Left; former communist), The Greens (Bündnis 90/Die Grünen; modern left), SPD (Social Democrats; traditional left), CDU/CSU (Christian Union; modern centrist-right), AfD (since 2013, Alternative für Deutschland; populist traditional-right). Data on the FDP (liberals), as a party alternating between left (1970s, in the chancellor Brandt and Schmidt eras) and right (1980s until 2021, in the Kohl and Merkel eras), were not collected separately, but are included in general left vs. non-left distinctions (see below). Sources: Scientific Service of the German Bundestag and state election administrations.
- *Left-right grouping of votes*: Votes for Die Linke, Grüne, and SPD (and Piratenpartei, another party of the left in 2011–2015) were summed and counted as votes for the (clearly) progressive-left camp. The remainder was added to the non-left, middle, bourgeois-conservative-right camp. As an alternative measure, we calculated a CDU−SPD contrast (votes for CDU/CSU minus votes for SPD). Union and SPD are the biggest and most important German parties and both are the prototypical parties of the right or the left. Data were collected for the period since 1959/1990 until 2019 or from 2000 until 2019.
- *Left-right categorization of state governments*: We developed for the governments (not electoral results) a 9 (right) to 1 (left) scale.[7] In nuances, a different gradation would be possible. To consider this we have also coded only the prime ministership (traditionally and in about 98 % of the cases 1 SPD, 2 CDU/CSU, 2011 and 2014 added 0.8 Grüne, 0.6 Linke). Period either from 1959 (or since 1990 in East Germany) or from 2000 until 2019.

A1.8 Data combination

Before the data were merged, the values within a survey and the same scale (e.g. IQB 2016 dimensions) were simply averaged, but between different years (e.g. IQB 2015 and 2016) or between different studies (e.g. IQB and PIRLS) the data were previously standardized to a uniform scale with the same mean and standard deviation (in the states common to the various studies) and only then averaged. This leads to more accurate and valid data overall. However, in order to be able to evaluate the size of effects in "natural units", if such meaningful scales exist (e.g. for student achievement), the mean values were transformed back into such a scale at the end.

A1.9 Analyses

Means, correlations, regressions, cross-sectional path models, and longitudinal path models comparing cross-lagged effects were calculated. Results will be presented as means, correlations and standardized regression coefficients (β). Path analyses are used to calculate the direct, indirect, net, and sum statistical effects of variables. In these analyses, the standardized path coefficients (β) between different variables are most important. Correlations are always added in parentheses. The differences between correlations and path coefficients help to quickly estimate the influence of other variables in a model (the larger the difference, the larger the statistical influence of other variables), and they enable model checking ($\Sigma r\beta = R^2 = 1$ – residual; the sum of the products of the correlations and beta coefficients gives the variance explained; residual/error is the unexplained variance) and a calculation of the proportion of variance explained by each single predictor ($R^2 = r\beta$). Missing paths correspond to small effects around zero. "Good" values for fit indices when models are not saturated are SRMR $\leq$ .08 or SRMR $\leq$ .05 and CFI $\geq$ .95 or CFI $\geq$ .97, and "acceptable" fit is achieved with SRMR $\leq$ .10 and CFI $\geq$ .95. All analyses were done in the naturally limited sample of 16 German states. No inferences from the sample to the population need to be tested (if at all possible by significance testing). (For further evaluation of significance tests, see: [116–118]). The basic problem of the analyses is the small state sample ($N = 16$). The following measures were taken to ensure the validity of the analyses and the interpretation of them:

- Checking the robustness of correlations across *different ability variables*: total score incl. Armed Forces, student achievement studies only, ability level of intellectual classes (95 %, top group students), ability level lower threshold (05 %, weak students), natives, migrants. The subgroups (95 %, 05 %, natives and migrants) were applied where appropriate.
- Testing the robustness of correlations across *different causal and effect variables* (standing for a common or similar phenomenon), e.g., operationalizing progressive vs. burgher education policies across different characteristics (such as central exit examinations, comprehensive schools), e.g., debt via two indicators, as private household debt and as state government debt.
- Use of *mean scales* composed of different items to increase the reliability and validity of measurement.
- Removal of *east-west* and *city-rural* (or urban-territorial) state differences in partial correlations (city states: Berlin, Bremen, Hamburg; all others rural-territorial states). It was theoretically examined whether removing these differences might conceal true relationships.

[7] 8 was assigned if CDU/CSU (Union) was governing alone; 7.2 Union and Freie Wähler (a smaller regional middle-right party, until now only in Bavaria); 7 Union and FDP (liberal party); 6.5 Union, FDP and Grüne (called "Jamaica"); 6 Union and SPD or Union and Grüne (Union prime minister); 5.5 Union, SPD and Grüne together (Union prime minister, called "Kenya coalition"); 5 SPD and Union (SPD prime minister); 4 SPD and FDP; 3 SPD alone or SPD, FDP and Grüne together (SPD prime minister) or Grüne and CDU (Grüne prime minister); 2 SPD and Grüne or SPD and Left; 1 Grüne and SPD (Grüne prime minister) or Left, SPD and Grüne (Left prime minister). A similar left-right classification of German parties was developed by Frerk in 2021 [115].

– Testing the robustness of the interplay of policy and ability over *different time periods* (e.g., from 1969 to 2000, 1973 to 2006, 2000 to 2016) and averaging the results of the different analyses.
– We also examined the policy-intelligence connection at the county level ($N = 83$ district recruiting offices of the German army).

The measures outlined increase the credibility of data and assumed relationships. For causal interpretations, it is also necessary to control alternative variables and to examine causal paths (e.g. from politics to educational policy decisions to test results). Finally, a convincing theory is necessary, supported as much as possible by studies on different data levels. Among explanatory causal factors, it is important to distinguish between explaining age differences (longitudinal), differences among students (individual differences), and differences among classes, schools, states, and nations (aggregate). Causes for individual development (e.g., amount of instruction) can deviate from causes for individual differences (e.g., doing *more* homework) or from causes for national differences (e.g., assigning *less* homework but more instruction by teachers). The meaning of variables may change with the level of data chosen (see above: ecological fallacy [49]). Methods and results must be used and understood in a reflective way according to the questions and data, this oriented to epistemic rationality.

A2 Results

A2.1 Means

Table A1
Means in cognitive ability studies (student achievement studies, psychometric IQ studies) on the level of German states

	CA mean (in IQ)	**Student mean** (in SASQ)	**Student m** (in school years)	**PISA mean** (00 03 06)	**PIRLS mean** (01 06)	**IQB mean** (09 11 12 15 16)	**Army IQ** (1998)
Baden-Wuertt.	100.91	505.65	9;9	510.50	554.91	500.77	100.85
Bavaria	102.41	515.59	10;0	520.39	558.38	515.95	102.11
Berlin	96.91	477.52	9;1	489.22	524.96	472.40	98.64
Brandenburg	97.92	487.93	9;4	487.24	533.54	496.81	96.94
Bremen	94.83	465.11	8;10	471.45	521.50	458.64	95.97
Hamburg	97.48	480.37	9;2	481.59	531.53	482.48	99.75
Hesse	99.15	491.84	9;5	492.48	542.63	492.76	100.85
Mecklenburg-P.	99.06	495.83	9;6	489.21	554.12	495.94	97.61
Lower Saxony	98.80	492.09	9;5	488.80	544.75	495.36	98.80
N. Rhine-W.	98.45	488.90	9;4	488.85	542.60	487.97	99.12
Rhineland-Pal.	99.85	499.15	9;7	496.71	555.16	496.41	99.59
Saarland	99.40	496.76	9;7	495.49	551.00	495.08	98.80
Saxony	101.64	512.09	9;11	514.36	557.25	513.38	100.35
Saxony-Anhalt	99.14	498.84	9;7	490.18	556.21	501.73	95.97
Schleswig-Hol.	99.55	494.72	9;6	493.81	544.75	497.67	101.07
Thuringia	100.94	507.56	9;10	503.61	565.58	502.67	99.68

Notes. Total cognitive ability score (CA) in IQ metric M=100 and SD=15 (student achievement and army/Bundeswehr studies combined, the latter one-fifth weighted), standardization adopted from student achievement studies; Student m: total student achievement studies quotient (SASQ) M=500 and SD=100 based on PISA, PIRLS, and IQB studies, scale adopted from these studies; PISA mean from 2000, 2003, and 2006; PIRLS mean from 2001 and 2006; IQB mean from 2009, 2011, 2012, 2015, and 2016; army (IQ test) from 1998, norms oriented to PISA 2000 but presented in IQ metric; school year equivalent in years and months based on student achievement studies (3rd column "in school years") assuming 42 SASQ points gain per school year and life year (student achievement study scale with M=500 and SD=100). Bavaria as the state with highest results in blue and Bremen with the lowest in red.

Table A2a
Means in student achievement studies (mean, 05 %-level, 95 %-level, level in Gymnasium, natives, immigrants) in SASQ (student achievement studies quotient)

	Student mean (in SASQ)	**05%-level** low ability	**95%-level** intellectual classes	**Gymnasium**	**Natives** no migration background	**Immigrants** with migration background
Baden-Wuertt.	505.65	330.11	660.60	590.27	530.04	464.03
Bavaria	515.59	342.71	667.58	599.58	539.71	484.50
Berlin	477.52	291.08	652.65	570.13	512.25	428.78
Brandenburg	487.93	320.66	660.46	574.36	495.94	464.15
Bremen	465.11	284.33	634.63	555.08	499.97	436.00
Hamburg	480.37	292.91	652.14	566.31	521.72	456.16
Hesse	491.84	316.02	652.27	572.35	518.52	451.41
Mecklenburg-P.	495.83	327.33	648.07	572.84	508.13	464.27
Lower Saxony	492.09	325.58	640.24	574.00	509.63	464.96
N. Rhine-W.	488.90	313.92	647.22	575.70	513.22	471.66
Rhineland-Pal.	499.15	324.18	653.58	584.31	523.10	469.23
Saarland	496.76	331.21	649.85	584.56	516.06	469.94
Saxony	512.09	356.89	667.11	593.12	527.21	479.09
Saxony-Anhalt	498.84	321.75	655.96	579.83	507.10	458.17
Schleswig-Hol.	494.72	322.01	653.35	587.42	515.03	483.62
Thuringia	507.56	342.34	655.24	579.66	519.11	449.09

Notes. Student m: total student achievement studies quotient (SASQ) M=500 and SD=100 based on PISA, PIRLS, and IQB studies, scale adopted from these studies; 05%-level: student ability level at 05% threshold, low ability; 95%-level: student ability level at 95% threshold, intellectual classes; Gymnasium: mean of students in Gymnasium; Natives' mean: mean of students without migration background, Migrants' mean: mean of students with migration background. All results in SASQ. Bavaria as the state with highest results in blue and Bremen with the lowest in red.

Table A2b
Means in student achievement studies (mean, 05 %-level, 95 %-level, level in Gymnasium, natives, immigrants) in IQ-metric

	Student mean (in IQ)	**05%-level** low ability	**95%-level** intellectual classes	**Gymnasium**	**Natives** no migration background	**Immigrants** with migration background
Baden-Wuertt.	100.85	73.67	123.25	112.70	103.66	93.76
Bavaria	102.34	75.56	124.29	114.09	105.11	96.83
Berlin	96.63	67.82	122.05	109.68	100.99	88.47
Brandenburg	98.19	72.25	123.23	110.31	98.55	93.78
Bremen	94.77	66.81	119.35	107.42	99.15	89.56
Hamburg	97.06	68.09	121.98	109.10	102.41	92.58
Hesse	98.78	71.56	122.00	110.01	101.93	91.87
Mecklenburg-P.	99.37	73.26	121.37	110.08	100.38	93.80
Lower Saxony	98.81	72.99	120.19	110.26	100.60	93.90
N. Rhine-W.	98.33	71.24	121.24	110.51	101.14	94.90
Rhineland-Pal.	99.87	72.78	122.19	111.80	102.62	94.54
Saarland	99.51	73.84	121.63	111.84	101.57	94.65
Saxony	101.81	77.69	124.22	113.12	103.24	96.02
Saxony-Anhalt	99.83	72.42	122.55	111.13	100.22	92.88
Schleswig-Hol.	99.21	72.46	122.16	112.27	101.41	96.70
Thuringia	101.13	75.51	122.44	111.10	102.02	91.52

Notes. See Table A2a, all the same but in IQ-metric (M=100, SD=15).

A2.2 Geography

Table A.3
Correlations with latitude and longitude

	Cognitive ability mean	Student mean
Latitude (north larger number)	−.66 [−.57]	−.65 [−.62]
Longitude (east larger number)	.11 [.27]	.15 [.35]

Notes. Correlations in brackets are partial correlations, in latitude west-east and city-rural state differences partialled out; in longitude the city-rural state difference. $N = 16$ German states.

We have a strong north-south divide in cognitive abilities within Germany (correlation with mean latitude of states $r = -.66$; see Table A3); i.e., in the south, young people are on average more intelligent, know more and have better student achievement. The results are stable across different dimensions (such as German and mathematics), across different studies (such as PISA and IQB), across different groups (students and young conscripts), and across different decades, but here with fluctuations due to the catch-up of the east. Internationally, a result with better values in the south is highly unusual. This is true of cross-country comparisons, such as countries in East Asia versus Southeast Asia, or North America versus Central America, or Europe versus Africa. Or within Europe, Scandinavia and Great Britain compared to Southern Europe, or Northeastern Europe compared to Southeastern Europe. Or within Italy, Spain and the USA comparing the regions and states of these countries: The north is always better than the south. The only known exception is Germany. And this is the case both in the west – Bavaria in the southwest, for example, has better results than Lower Saxony in the northwest – and in the east – Saxony, for example, performs better than Mecklenburg-Vorpommern. However, the south-north effect is purely descriptive; latitude itself is not a cause. The magnitude, globality, and stability of the differences suggest stable and generally effective causal factors related to geography.

A2.3 Correlations, possible causes and consequences: Background factor genes

At the international level, evolutionary genetic factors, together with culture, represent the two background factors that can be used to explain international differences in intelligence and knowledge statistically very well, also indirectly via an effect on education [29,119].

One evidence at the international level is a correlation with haplogroups: The same haplogroups that correlate with differences in a country sample from Europe, the Middle East, and North Africa (MEA) also correlate with regional differences in Italy and Spain. What about differences within Germany? Unfortunately, there are no data for federal states, but at least for the regions North, South, West and East Germany.

Table A.4
Genes (haplogroups)

	PISA 2006 mean
Haplogroups A, internationally positive	−.72
Haplogroups B, internationally negative	+.76

Notes. Data and results from Rindermann, Woodley & Stratford from 2012 [45]; haplogroups A represent cultural progress in the Neolithic (development of symbolic communication and art), haplogroups B in the Paleolithic (agriculture and livestock); $N = 4$ North, South, West, East Germany regions.

Known genetic indicators of cognitive evolution do not correlate in the expected direction in Germany, but exactly the reverse (Table A4). This does not mean that there are no genetic factors at all, but they seem to have nothing to do with global evolutionary processes. An example of such more local developments from the UK: Genes representing lower education are more common in economically less modern developed regions such as coal mining areas [72]. People that are more intelligent were more likely to leave such regions.

A2.4 Scatterplots

Figure A1 shows a scatter plot of votes burgher-civic-conservative-right (x-axis) with the cognitive ability mean (y-axis; $r = .82$, $N = 16$ states; Table 11). As can be seen, there are two outliers, Bremen in the lower left and Brandenburg in the middle left.

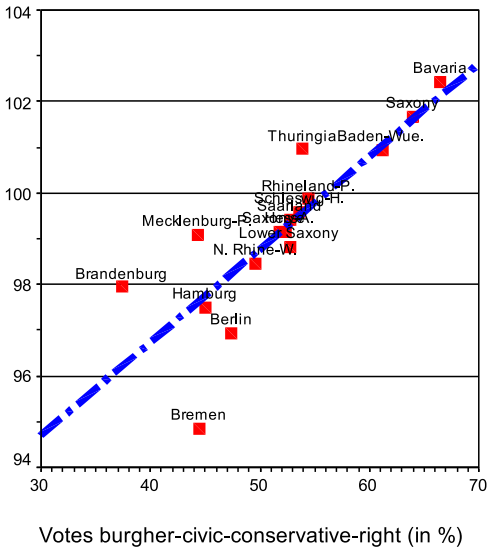


Fig. A1. Scatter plot of votes burgher-civic-conservative-right (x-axis) with the cognitive ability mean (y-axis; $r = .82$, $N = 16$ states).

When Bremen is removed from the data set, the correlation increases to $r = .86$ ($N = 15$); when Brandenburg is additionally removed, the correlation is $r = .89$ ($N = 14$). As indicated by the partial correlation (removing west-east and city-rural state differences; $r_p = .88$) the "true" correlation between politics and cognitive ability at the state level is maybe even higher.

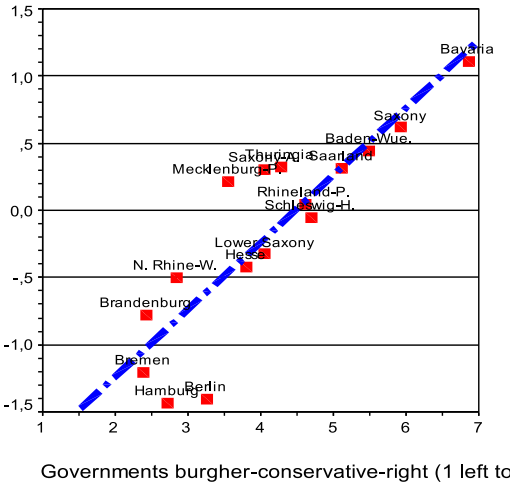


Fig. A2. Scatter plot of governments burgher-conservative-right (x-axis) with the conservative educational policies mean (y-axis; $r = .87$, $N = 16$ states).

Figure A2 shows a scatter plot of governments burgher-conservative-right (x-axis) with the conservative educational policies mean (y-axis, scale based on *z*-standardizations; $r = .87$, $N = 16$ states; Table 12). As can be seen, there are three outliers, Bavaria at the top right and Berlin and Hamburg at the bottom. When Bavaria is removed from the data set, the correlation decreases to $r = .83$ ($N = 15$); when Berlin and Hamburg are additionally removed, the correlation is $r = .86$ ($N = 13$). Outliers do not bias the positive correlation between governments and education policies.

References

[1] J. Baumert, R. Watermann, Institutionelle und regionale Variabilität und die Sicherung gemeinsamer Standards in der gymnasialen Oberstufe [Institutional and regional variability and keeping up standards in upper secondary schools], in: J. Baumert, W. Bos, R. Lehmann (Eds.), TIMSS/III, Opladen: Leske and Budrich), 2000, pp. 317–372.

[2] H.J. Ebenrett, D. Hansen, K.J. Puzicha, Verlust von Humankapital in Regionen mit hoher Arbeitslosigkeit [Loss of human capital in regions of high unemployment], Aus Politik und Zeitgeschichte B6– 7 (2003) 25–31.
[3] E.A. Hanushek, J. Ruhose, L. Woessmann, Economic gains from educational reform by US states, J. Hum. Cap. 11 (2017) 447–486, https://doi.org/10.1086/694454.
[4] S. Kanazawa, IQ and the wealth of states, Intelligence 34 (2006) 593–600, https://doi.org/10.1016/j.intell.2006.04.003.
[5] M.A. McDaniel, Estimating state IQ: measurement challenges and preliminary correlates, Intelligence 34 (2006) 607–619, https://doi.org/10.1016/j.intell.2006.08.007.
[6] R. Lynn, J. Fuerst, E.O.W. Kirkegaard, Regional differences in intelligence in 22 countries and their economic, social and demographic correlates: a review, Intelligence 69 (2018) 24–36, https://doi.org/10.1016/j.intell.2018.04.004.
[7] E.A. Hanushek, L. Woessmann, The Knowledge Capital of Nations. Education and the Economics of Growth, MIT Press, Cambridge, 2015.
[8] G. Jones, Hive Mind: How Your Nation's IQ Matters So Much More than Your Own, Stanford University Press, 2016, https://doi.org/10.1080/15582159.2017.1345242.
[9] R. Lynn, T. Vanhanen, IQ and the Wealth of Nations, Praeger, 2002.
[10] E. Roivainen, Economic, educational, and IQ gains in eastern Germany 1990-2006, Intelligence 40 (6) (2012) 571–575, https://doi.org/10.1016/j.intell.2012.07.007.
[11] J. Kraus, Fünfzehn Fragen und Antworten zur innerdeutschen PISA-Studie [Fifteen questions and answers on the German regional PISA study], in: H. Schmoll, J. Kraus, J.-D. Gauger, H. Grewe (Eds.), PISA-E und was nun? Bilanz des innerdeutschen Schulvergleichs [PISA-E and what now? Taking stock of the intra-German school comparison], Konrad-Adenauer-Stiftung, 2002, pp. 22–49.
[12] H. Schmoll, Niederschmetternder befund [devastating finding], Frankfurter Allgemeine Zeitung 132 (2014) 10, 10.06.14.
[13] L. Stankov, Conservatism and cognitive ability, Intelligence 37 (2009) 294–304, https://doi.org/10.1016/j.intell.2008.12.007.
[14] S.A. Saribay, O. Yilmaz, Analytic cognitive style and cognitive ability differentially predict religiosity and social conservatism, Pers. Indiv. Differ. 114 (2017) 24–29, https://doi.org/10.1016/j.paid.2017.03.056.
[15] E. Onraet, A. Van Hiel, K. Dhont, G. Hodson, M. Schittekatte, S. De Pauw, The association of cognitive ability with right-wing ideological attitudes and prejudice: a meta-analytic review, Eur. J. Pers. 29 (6) (2015) 599–621, https://doi.org/10.1002/per.2027.
[16] N. Carl, Cognitive ability and party identity in the United States, Intelligence 47 (2014) 3–9, https://doi.org/10.1016/j.intell.2014.08.003.
[17] N. Carl, IQ and political attitudes across British regions and local authorities, Intelligence 69 (2018) 169–175, https://doi.org/10.1016/j.intell.2018.06.002.
[18] H. Rindermann, C. Flores-Mendoza, M.A. Woodley, Political orientations, intelligence and education, Intelligence 40 (2) (2012) 217–225, https://doi.org/10.1016/j.intell.2011.11.005.
[19] S. Feldman, Ch Johnston, Understanding the determinants of political ideology: implications of structural complexity, Polit. Psychol. 35 (3) (2014) 337–358, https://doi.org/10.1111/pops.12055.
[20] B. Caplan, S.C. Miller, Intelligence makes people think like economists: evidence from the general social survey, Intelligence 38 (6) (2010) 636–647, https://doi.org/10.1016/j.intell.2010.09.005.
[21] World Values Survey, Data from 1981 to 2014, WVS_Longitudinal_1981-2014_spss_v_2014_06_17_Beta, 2014. Retrieved from, www.worldvaluessurvey.org.
[22] Y. Ganzach, Intelligence and the rationality of political preferences, Intelligence 69 (2018) 59–70, https://doi.org/10.1016/j.intell.2018.05.002.
[23] L.S. Gottfredson, g, jobs and life, in: H. Nyborg (Ed.), The Scientific Study of General Intelligence. Tribute to Arthur R. Jensen, Pergamon, 2003, pp. 293–342.
[24] S. Pinker, Rationality: what it is, why it seems scarce, why it matters, Viking (2021).
[25] M. Weber, The Protestant Ethic and the Spirit of Capitalism, Routledge, 2001/1905.
[26] L.E. Harrison, S.P. Huntington, Culture Matters: How Values Shape Human Progress, Basic Books, 2000.
[27] S.O. Becker, L. Woessmann, Was Weber wrong? A human capital theory of Protestant economic history, Q. J. Econ. 124 (2) (2009) 531–596, https://doi.org/10.1162/qjec.2009.124.2.531.
[28] M.R. Platas, Culture and the Persistence of Educational Inequality: Lessons from the Muslim-Christian Education Gap in Africa, New York University Abu Dhabi, Abu Dhabi, 2018. Retrieved from, https://mplatas.files.wordpress.com/2018/05/persistence_may2018.pdf.
[29] H. Rindermann, Cognitive Capitalism: Human Capital and the Wellbeing of Nations, Cambridge University Press, 2018.
[30] H. Rindermann, D. Becker, Flynn-effect and economic growth: do national increases in intelligence lead to increases in GDP? Intelligence 69 (2018) 87–93, https://doi.org/10.1016/j.intell.2018.05.001.
[31] J. Kramer, Allgemeine Intelligenz and beruflicher Erfolg in Deutschland [General mental ability and occupational success in Germany], Psychol. Rundsch. 60 (2009) 82–98, https://doi.org/10.1026/0033-3042.60.2.82.
[32] P. Irwing, R. Lynn, The relation between childhood IQ and income in middle age, J. Soc. Polit. Econ. Stud. 31 (2) (2006) 191–196.
[33] U.R. Hülsheger, G.W. Maier, T. Stumpp, Validity of general mental ability for the prediction of job performance and training success in Germany: a meta-analysis, Int. J. Sel. Assess. 15 (1) (2007) 3–18, https://doi.org/10.1111/j.1468-2389.2007.00363.x.
[34] F.L. Schmidt, J.E. Hunter, General mental ability in the world of work: occupational attainment and job performance, J. Pers. Soc. Psychol. 86 (1) (2004) 162–173, https://doi.org/10.1037/0022-3514.86.1.162.
[35] H.J. Ebenrett, D. Hansen, K.J. Puzicha, "Brain drain" in deutschen Regionen [Brain drain in German regions], Report of the Psychologische Dienst Bundeswehr (1/02) (2002).
[36] M. Jahoda, P.F. Lazarsfeld, H. Zeisel, Marienthal: the Sociography of an Unemployed Community, Routledge, 2002/1933.
[37] R. Block, K. Klemm, Gleichwertige Lebensverhältnisse im Bundesgebiet? [Equivalent living conditions in Germany? Universität Duisburg-Essen (2005).
[38] K. Klemm, Pisa funktioniert nicht wie die Bundesliga, 2010, May 17 [Pisa does not work like the German football league.] Süddeutsche Zeitung. Retrieved from, www.sueddeutsche.de/karriere/bildungsforscher-klemm-pisa-funktioniert-nicht-wie-die-bundesliga-1.555498.
[39] P. Stanat, Schulleistungen von Jugendlichen mit Migrationshintergrund: die Rolle der Zusammensetzung der Schülerschaft [Student achievement of immigrant youth: the role of student body composition], in: J. Baumert, P. Stanat, R. Watermann (Eds.), Herkunftsbedingte Disparitäten im Bildungswesen [Origin-related disparities in education], Springer VS, 2006, pp. 189–219.
[40] P. Stanat, Kompetenzen von Viertklässlern [Competences of fourth graders]. Der Spiegel, 2017, October 13. Retrieved from, www.spiegel.de/lebenundlernen/schule/iqb-bildungsstudie-die-kinder-haben-keine-schuld-a-1172835.html.
[41] J.J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, D. Cesarini, Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals, Nat. Genet. 50 (8) (2018) 1112–1121, https://doi.org/10.1038/s41588-018-0147-3.
[42] S.v. Stumm, R. Plomin, Using DNA to predict intelligence, Intelligence 86 (2021) 101530, https://doi.org/10.1016/j.intell.2021.101530.
[43] Y. Kovas, S. Malykh, D. Gaysina (Eds.), Behavioural Genetics for Education, Palgrave Macmillan UK, London, 2016, 1057/9781137437327.
[44] D. Piffer, A review of intelligence GWAS hits: their relationship to country IQ and the issue of spatial autocorrelation, Intelligence 53 (2015) 43–50, https://doi.org/10.1016/j.intell.2015.08.008.
[45] H. Rindermann, M.A. Woodley, J. Stratford, Haplogroups as evolutionary markers of cognitive ability, Intelligence 40 (2012) 362–375, https://doi.org/10.1016/j.intell.2012.04.002.
[46] M. Steppan, Protestantism and intelligence, Int. J. Educ. Psychol. Assess. 5 (2010) 134–154.
[47] S.O. Becker, L. Woessmann, How Luther's quest for education changed German economic history: 9+5 theses on the effects of the Protestant Reformation, in: J.-P. Carvalho, S. Iyer, J. Rubin (Eds.), Advances in the Economics of Religion, Palgrave Macmillan, 2019, pp. 215–227.
[48] I.J. Deary, G.D. Batty, C.R. Gale, Childhood intelligence predicts voter turnout, voting preferences, and political involvement in adulthood: the 1970 British Cohort Study, Intelligence 36 (2008) 548–555, https://doi.org/10.1016/j.intell.2008.09.001.
[49] A.J. Idrovo, Three criteria for ecological fallacy, Environ. Health Perspect. 119 (8) (2011) A332, https://doi.org/10.1289/ehp.1103768.
[50] O. Köller, M. Knigge, B. Tesch, Sprachliche Kompetenzen im Ländervergleich [Language competencies in state comparison], Waxmann, 2010.

[51] H. Rindermann, Q.S.N. Hoang, A.E.E. Baumeister, Cognitive ability, parenting and instruction in Vietnam and Germany, Intelligence 41 (2013) 366–377, https://doi.org/10.1016/j.intell.2013.05.011.
[52] J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, M. Weiß, Pisa 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland [PISA 2000 – A differentiated look at the states of the Federal Republic of Germany, 2003 (Opladen: Leske + Budrich).
[53] D. Flor, K. Ingenkamp, W.H. Schreiber, Schulleistungsvergleiche zwischen Bundesländern [Student achievement comparisons between German states], Beltz, 1992.
[54] G. Trost, L. Pauels, B. Schneider, Repräsentativerhebung an deutschen Abiturienten [Representative survey of German high school graduates]. Bad Godesberg: Studienstiftung des Deutschen Volkes, 1976.
[55] B.-P. Robra, H. Schmitt, Die Ergebnisse der IMPP-Prüfungen sind keine geeignete Grundlage für eine jährliche leistungsbezogene Mittelzuweisung an deutschen medizinischen Fakultäten [Results of the IMPP exams are not a suitable basis for an annual performance-orientated allocation of funds at German medical faculties], Med. Ausbild. 18 (2001) 159–164.
[56] J. Komlos, P. Kriwy, The biological standard of living in the two Germanies, Ger. Econ. Rev. 4 (2003) 493–507, https://doi.org/10.1111/j.1465-6485.2003.00089.x.
[57] S.J. Ritchie, E.M. Tucker-Drob, How much does education improve intelligence? A meta-analysis, Psychol. Sci. 29 (2018) 1358–1369, https://doi.org/10.1177/0956797618774253.
[58] S. Bergold, L. Wirthwein, D.H. Rost, R. Steinmayr, What happens if the same curriculum is taught in five instead of six years? A quasi-experimental investigation of the effect of schooling on intelligence, Cognit. Dev. 44 (2017) 98–109, https://doi.org/10.1016/j.cogdev.2017.08.012.
[59] K.A. Heller, Ch Perleth, Kognitiver Fähigkeits-Test (KFT 4-12+R) [Cognitive Abilities Test (CogAT 4-12+R)], 2000. Beltz.
[60] F.K. Ringer, The Decline of the German Mandarins, Harvard University Press, 1969.
[61] J.H. Bishop, The effect of national standards and curriculum-based exams on achievement, Am. Econ. Rev. 87 (1997) 260–264.
[62] F. Büchel, H. Jürges, K. Schneider, Die Auswirkungen zentraler Abschlussprüfungen auf die Schulleistung – quasi-experimentelle Befunde aus der deutschen TIMSS-Stichprobe [The impact of central examinations on student achievement – quasi-experimental findings from the German TIMSS sample], Vierteljahrshefte Wirtschaftsforsch. 72 (2003) 238–251, https://doi.org/10.3790/vjh.72.2.238.
[63] K. Koerselman, Incentives from curriculum tracking, Econ. Educ. Rev. 32 (2013) 140–150, https://doi.org/10.1016/j.econedurev.2012.08.003.
[64] J.A.C. Hattie, Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement, 2009, https://doi.org/10.4324/9780203887332. Routledge.
[65] H.W. Stevenson, J.W. Stigler, The Learning Gap. Why Our Schools Are Failing and what We Can Learn from Japanese and Chinese Education, Summit, 1987.
[66] N.A.S. Burhan, M.R. Mohamad, Y. Kurniawan, A.H. Sidek, The impact of low, average, and high IQ on economic growth and technological progress: do all individuals contribute equally? Intelligence 46 (2014) 1–8, https://doi.org/10.1016/j.intell.2014.04.006.
[67] Th R. Coyle, H. Rindermann, D.G. Hancock, J. Freeman, Nonlinear effects of cognitive ability on economic productivity, J. Indiv. Differ. 39 (1) (2018) 39–47, https://doi.org/10.1027/1614-0001/a000247.
[68] M. Levels, J. Dronkers, Educational performance of native and immigrant children from various countries of origin, Ethn. Racial Stud. 31 (2008) 1404–1425, https://doi.org/10.1080/01419870701682238.
[69] H. Rindermann, J. Thompson, The cognitive competences of immigrant and native students across the world, J. Biosoc. Sci. 48 (2016) 66–93, https://doi.org/10.1017/S0021932014000480.
[70] OECD, Settling in 2018: Indicators of Immigrant Integration, Paris: OECD, European Union, 2018.
[71] D. McClendon, Muslims in Sub-saharan Africa Are Twice as Likely as Christians to Have No Formal Education, in: Pew Research Center, 14 December, 2016. Retrieved from, www.pewresearch.org/fact-tank/2016/12/14/muslims-in-sub-saharan-africa-are-twice-as-likely-as-christians-to-have-no-formal-education.
[72] A. Abdellaoui, D. Hugh-Jones, L. Yengo, K.E. Kemper, M.G. Nivard, L. Veul, P.M. Visscher, Genetic correlates of social stratification in Great Britain, Nat. Human Behav. 3 (12) (2019) 1332–1342, https://doi.org/10.1038/s41562-019-0757-5.
[73] A.R. Luria, Cognitive Development. Its Cultural and Social Foundations, Harvard University Press, 1976/1974.
[74] R. Proctor, L.L. Schiebinger (Eds.), Agnotology: the Making and Unmaking of Ignorance, Stanford University Press, Stanford, 2008.
[75] R.B. Kline, Principles and Practice of Structural Equation Modeling, Guilford Press, 2016.
[76] J.L. Duarte, J.T. Crawford, C. Stern, J. Haidt, L. Jussim, P.E. Tetlock, Political diversity will improve social psychological science, Behav. Brain Sci. 38 (1–13) (2015) 45–58, https://doi.org/10.1017/S0140525X14000430.
[77] S.J. Ceci, W.M. Williams, Socio-political values infiltrate the assessment of scientific research, in: J.T. Crawford, L. Jussim (Eds.), Politics of Social Psychology, Routledge, Psychology Press, 2018, pp. 156–167.
[78] C.L. Frisby, R. Redding, W.T. O'Donohue, S.O. Lilienfeld (Eds.), Political Bias in Psychology: Nature, Scope, and Solutions, Springer, 2023, https://doi.org/10.1007/978-3-031-29148-7_1.
[79] G. Geher, Politics and academic values in higher education. Interview by the Dissenter, 15.10.2021, 2021. Retrieved from, www.youtube.com/watch?v=Q3XVnbQBw0A.
[80] Y. Inbar, J. Lammers, Political diversity in social and personality psychology, Perspect. Psychol. Sci. 7 (2012) 496–503, https://doi.org/10.1177/1745691612448792.
[81] U. Peters, N. Honeycutt, A. De Block, L. Jussim, Ideological diversity, hostility, and discrimination in philosophy, Phil. Psychol. 33 (2020) 511–548, https://doi.org/10.1080/09515089.2020.1743257.
[82] N. Steindl, C. Lauerer, Th Hanitzsch, Journalismus in Deutschland: aktuelle Befunde zu Kontinuität und Wandel im deutschen Journalismus [Journalism in Germany. Current findings on continuity and change in German journalism], Publizistik 62 (2017) 401–423, https://doi.org/10.1007/s11616-017-0378-9.
[83] S. Weischenberg, M. Malik, A. Scholl, Die Souffleure der Mediengesellschaft. Report über die Journalisten in Deutschland [The prompters of the media society, UVK, 2006. Report on journalists in Germany].
[84] S. Greenstein, F. Zhu, Do experts or collective intelligence write with more bias? Evidence from Encyclopedia Britannica and Wikipedia, MIS Q. 42 (2018) 945–959.
[85] S. Tezuka, L. Ashtear, The left-wing bias of Wikipedia. Is Wikipedia's neutral point of view truly dead? Critic (2020), 22 October 2020, https://thecritic.co.uk/the-left-wing-bias-of-wikipedia.
[86] N. Thompson, D. Hanley, Science is shaped by Wikipedia: evidence from a randomized control trial, SSRN Electron. J. (2017), https://doi.org/10.2139/ssrn.3039505.
[87] J. Kempner, The chilling effect: how do researchers react to controversy? PLoS Med. 5 (11) (2008) e222, 1371/journal.pmed.0050222.
[88] J.-W. Laliberté, Long-term contextual effects in education: schools and neighborhoods, Am. Econ. J. Econ. Pol. 13 (2) (2021) 336–377, https://doi.org/10.1257/pol.20190257.
[89] N.G. Shakeshaft, M. Trzaskowski, A. McMillan, K. Rimfeld, E. Krapohl, C.M.A. Haworth, R. Plomin, Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16, PLoS One 8 (12) (2013) e80341, https://doi.org/10.1371/journal.pone.0080341.
[90] S.B. Kaufman, M.R. Reynolds, X. Liu, A.S. Kaufman, K.S. McGrew, Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests, Intelligence 40 (2012) 123–138, https://doi.org/10.1016/j.intell.2012.01.009.
[91] H. Rindermann, A.E.E. Baumeister, Validating the interpretations of PISA and TIMSS tasks: a rating study, Int. J. Test. 15 (1) (2015) 1–22, https://doi.org/10.1080/15305058.2014.966911.
[92] A. Pokropek, G.N. Marks, F. Borgonovi, How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? J. Educ. Psychol. (2021) https://doi.org/10.1037/edu0000687.
[93] M. Brunner, No g in education? Learn. Indiv Differ 18 (2) (2008) 152–165, https://doi.org/10.1016/j.lindif.2007.08.005.
[94] PISA-Konsortium Deutschland, PISA 2006 in Deutschland. Die Kompetenzen der Jugendlichen im dritten Ländervergleich, Waxmann, 2008 [PISA 2006 in Germany. The competences of young people in the third state comparison].
[95] W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, R. Valtin, IGLU-E 2006 [PIRLS 2006], Waxmann, 2008.

[96] P. Stanat, S. Schipolowski, C. Rjosk, S. Weirich, N. Haag, Hrsg, IQB-bildungstrend 2016, Waxmann, 2017 [IQB education trend].
[97] Destatis, Kindertagesbetreuung [Daycare], Retrieved from, www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Soziales/Sozialleistungen/Kindertagesbetreuung/Tabellen/Tabellen_Betreuungsquote.html, 2018.
[98] Spiegel, Chaos Schule, Mißwirtschaft im Klassenzimmer [Chaos school, mismanagement in the classroom], Spiegel 45 (41) (1991, October 7) 48–73.
[99] H. Diefenbach, M. Klein, Bringing boys back in, Z. für Padagogik 48 (2002) 938–958, https://doi.org/10.25656/01:3868.
[100] T. Tricarico, Zensuren führen zu Frust oder Beschämung [Grades lead to frustration or embarrassment]. Die Welt, Retrieved from, www.welt.de/wirtschaft/karriere/bildung/article136969526/Zensuren-fuehren-zu-Frust-oder-Beschaemung.html, 2015, January 31.
[101] Ch Böhm, Evaluation der Pädagogik Wolfgang Klafki [Evaluation of Wolfgang Klafki's pedagogy], Kovac, 2008.
[102] S.v. Below, Bildungssysteme und soziale Ungleichheit [Education systems and social inequality], Leske+Budrich, 2002.
[103] H. Fend, Die Pädagogik des Neokonservatismus [The pedagogy of neoconservatism], Suhrkamp, 1984.
[104] F.E. Weinert, Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit [Comparative achievement measurement in schools – a controversial matter of course], in: F.E. Weinert (Ed.), Leistungsmessungen in Schulen [Achievement Measurement in Schools], Beltz, 2001, pp. 17–31.
[105] CHE, CHE-forschungsranking 2009, 2009 [CHE ranking of research]. Gütersloh: Working Paper 130.
[106] Wissenschaftsrat, Entwicklung der Fachstudiendauer an Universitäten von 1999 bis 2003 [Development of the duration of study at universities from 1999 to 2003], Bericht des Wissenschaftsrates, Berlin, 2005.
[107] Destatis, Hochschulen auf einen Blick [Universities at a glance], Statistisches Bundesamt, Wiesbaden, 2018 (Destatis).
[108] Institut der deutschen Wirtschaft, Deutschland in Zahlen, 2019 [Germany by the numbers.]. Retrieved from deutschlandinzahlen.de.
[109] Creditreform, Schuldneratlas Deutschland [debtor atlas Germany], Neuss: Credit (2018).
[110] Schufa (Schutzgemeinschaft für allgemeine Kreditsicherung), SCHUFA Kredit-Kompass 2010 [SCHUFA Credit Compass 2010], Schufa, Wiesbaden, 2010.
[111] Destatis, Statistisches Bundesamt, Retrieved from, www.destatis.de, 2018.
[112] Bertelsmann-Stiftung, Kommunale Sozialausgaben – wie der Bund sinnvoll helfen kann [Municipal social spending – how the federal government can help in a meaningful way], Bertelsmann (2015).
[113] Bundeskriminalamt (BKA, PKS Polizeiliche Kriminalstatistik 2017 [PKS Polizeiliche Kriminalstatistik 2017], BKA, Wiesbaden, 2017.
[114] Statista, Religionszugehörigkeit der Deutschen nach Bundesländern im Jahr 2011 [Religious affiliation of Germans by federal state in 2011], Retrieved from, https://de.statista.com/statistik/daten/studie/201622/umfrage/religionszugehoerigkeit-der-deutschen-nach-bundeslaendern, 2018.
[115] C. Frerk, Parteipräferenzen. Profile [Party preferences. Profiles], Retrieved from website: https://fowid.de/meldung/parteipraeferenzen-profile, 2021.
[116] J.E. Hunter, Needed: a ban on the significance test, Psychol. Sci. 8 (1) (1997) 3–7, https://doi.org/10.1111/j.1467-9280.1997.tb00534.x.
[117] B.B. McShane, D. Gal, A. Gelman, C. Robert, J.L. Tackett, Abandon statistical significance, Am. Statistician 73 (sup1) (2019) 235–245, https://doi.org/10.1080/00031305.2018.1527253.
[118] R.L. Wasserstein, A.L. Schirm, N.A. Lazar, Moving to a world beyond "p < 0.05", Am. Statistician 73 (1) (2019) 1–19, https://doi.org/10.1080/00031305.2019.1583913.
[119] D. Becker, H. Rindermann, The relationship between cross-national genetic distances and IQ-differences, Pers. Indiv. Differ. 98 (2016) 300–310, https://doi.org/10.1016/j.paid.2016.03.050.