Project Report for Data Literacy 2023/24
Grade Inflation in the German School System - Causes and Effects

Jonathan Schwab [*1] **Lars Kasüschke** [*2] **Niklas Munkes** [*3] **Tom Freudenmann** [*4]

Abstract

Introduction

The Abitur grades have constantly increased in the German school system over the past decades. Every year, when the Abitur takes place, the grades and the difficulty of the exercises are extensively discussed in the media and have been part of a fierce research discussion for decades. The central focus of the discourse revolves around the question of whether grade inflation occurs, signifying a rise in grades without a corresponding increase in competence or knowledge.

The discourse has predominantly centred around mathematics, since the difficulty of exercises is easiest to compare. The line is drawn between mathematicians arguing that specific exercises are easier than exercises in the past (Kühnel, 2015) (Jahnke et al., 2014) (Lemmermeyer et al., 2019) and studies claiming that grade inflation cannot be reliably proven since the competence of students has also increased (Schleithoff, 2015). In 2015, a data-driven approach was employed, involving the analysis of comprehensive data on the education system. The analysis was promising, but not yet enough to neglect the claim of a grade inflation (Grözinger & Baillet, 2015).

This paper expands on that work, disproving the claim that grade inflation is the main cause of the observed trend. An explanatory framework for the improvement of Abitur grades is provided by this data analysis, building upon past research. All analysed data is taken from official federal resources such as the German Federal Statistics Office.

[*]Equal contribution [1]Matrikelnummer 6765897, jonathan.schwab@student.uni-tuebingen.de, MSc Computer Science [2]Matrikelnummer 4247775, lars.kasueschke@student.uni-tuebingen.de, BSc Computer Science [3]Matrikelnummer 4269436, niklas.munkes@student.uni-tuebingen.de, MSc Media Informatics [4]Matrikelnummer 6631101, tom.freudenmann@student.uni-tuebingen.de, MSc Computer Science.

Project report for the "Data Literacy" course at the University of Tübingen, Winter 2023/24 (Module ML4201). Style template based on the ICML style files 2023.

The study undertakes an analysis of the quantifiable impacts stemming from the upward trend in Abitur grades within the educational system. Prognostications are offered concerning the trajectory of future grade developments and the implications for the German education system.

Methods

The Pearson correlation coefficient (Rodgers & Nicewander, 1988), denoted as r, is a statistical measure used to assess the linear relationship between two sets of data, X and Y. It is computed as the ratio of the sample covariance of the X and Y to the product of their sample standard deviations:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (1)$$

Datasets: The datasets used in this paper can be grouped into causes and effects. Therefore, the causes are societal, demographic, or political factors influencing the quality of the German school system. This could be the number of students, teachers, or budget provided by the German government. In contrast, this paper defines the effects of the causes as observable measures of the students' performance. Examples are the average grades, PISA study results, or the rate of repeaters and school-leavers.

The first cause dataset is provided in the *Fachreport Schuljahr 2020/21* of the Statistische Bundesamt and contains the number of teachers from 1992 until 2020. The dataset groups them primarily according to their contract type, federal state, and school type. This paper merges the teacher counts with two student datasets, which are published in the *Genesis* database of the Statistische Bundesamt. Both provide the number of children as different groupings and aggregations. The first contains the number of children per grade and school type for the years 1998 to 2022. In contrast, the second table provides the absolute amount of children, leavers, and beginners in each federal state from 1997 to 2022. Therefore, the analysis of the merged dataset can only be conducted separately for school types and federal states.

To conclude the causes, this paper introduces the budget per child, which is provided in the *Genesis* database of the Statis-

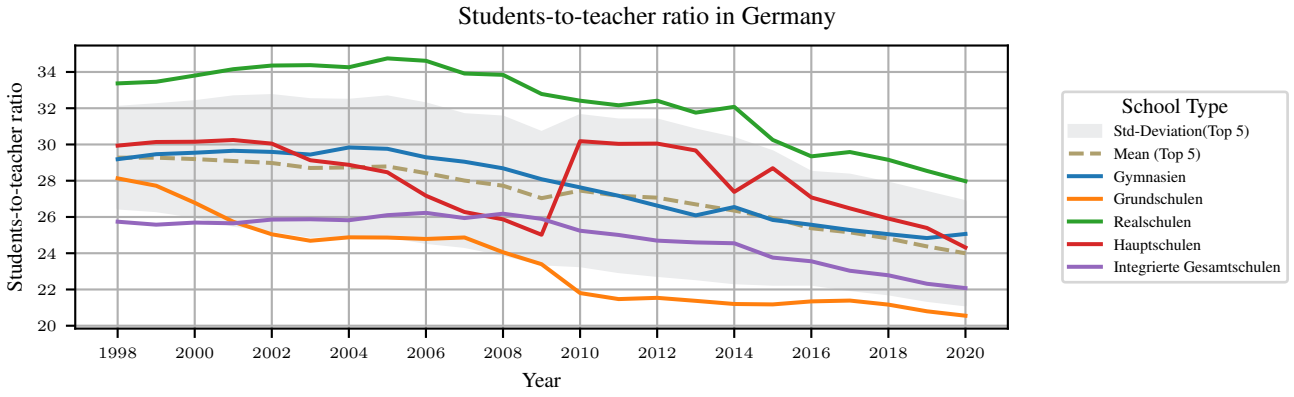

Figure 1. Students-to-teacher ratio of the five most common school types in Germany. The ratio of full- and part-time teachers is displayed foreach school type and aggregated to their mean (▬ ▬) and standard deviation (▬).

tische Bundesamt. The dataset contains the budget per child for the years 2010 to 2022 and is grouped by federal states. In contrast to the demographic and societal causes above, the budget models a direct political factor. To adjust the budget to inflation, it is multiplied with the *Verbraucherpreisindex* relative to 2022 provided by the Statistische Bundesamt.

Moreover, the primary effects on students' performance are the basis for the analysis of the German school system. One of the few publicly available datasets containing grades is the average Abitur grades per federal state. The grades are published every year by the Kultusminister Konferenz. Each file contains the count of children per written grade and federal state. In addition, the grades are defined in 0.1 steps, with 4.0 as the worst and 1.0 as the best grade. Furthermore, the amount of children who failed with a grade worse than 4.0 is aggregated in an additional column.

Although this is a great model for the performance of children attending grammar schools, a general performance measure for all school types is required. Accordingly, this paper uses the number of repeaters derived from the *GENESIS* Database of the Statistische Bundesamt. There, the absolute count of repeaters by federal state, school type, and year is provided for the years 1998 to 2022.

First trends and basic analysis Having introduced all used datasets, this paragraph aims to investigate potential patterns through an exploratory data analysis of the potential causes and effects.

Firstly, regard the demographic effects on the number of children attending school and teachers employed by school type and federal state. The exploratory data analysis has shown that the number of schoolchildren decreases steadily from 1998 to 2014. Instead, it increased from 2019 to 2022 because more children started their education and fewer left school. Furthermore, more children graduate from grammar schools with university entrance qualifications. This demographic effect is combined with an increasing number of teachers across all German school types and federal states. Although, the percentage of part-time teachers is increasing, the number of full-time teachers is decreasing until 2020.

Given the hypothesis that having more teachers per student increases the quality of teaching, the datasets can be merged. As already explained, this merge can only been done separately for school types and federal states. Furthermore, the student-to-teacher ratio is calculated over all full- and part-time teachers, since they represent the majority of the distribution. In contrast, the teachers who are employed on an hourly basis are excluded due to their insignificant impact on the teaching quality and sparse representation in the data. The results (Figure 1) show that from 1998 to 2020, the ratio decreased for the five most common school types. As a result, the average decreases from 29 to 24 children per teacher. Together with the hypothesis, it follows that the quality of teaching should increase, and thus the performance measures should increase.

Besides the demographic measures, the federal states plan their school budgets independently. The analysis of the adjusted budget to inflation per child has shown that it steadily increases for all federal states. Although, this may be caused by the increasing number of teachers and the goals of digitalization at schools (TODO source).

Now that the basic effects that may influence the students' performance have been identified, it is possible to study the performance measures. As the analysis of the students datasets has shown, more children are attending grammar schools in Germany. Thus, the average Abitur grade of the children is a great measure of the performance of many children. Figure 2 shows that the average grades are increasing in all federal states. Furthermore, a linear regression can be employed to represent their mean. Importantly, the regression is calculated on the data before 2021 because of the COVID-19 pandemic beginning in 2020. In 2022, the

grades significantly increased compared to the years before the pandemic. This could indicate that the pandemic has had novel consequences for the educational system. Due to the lack of data following the pandemic, this paper will solely focus on the linear trend until 2020. Furthermore, an additional analysis of the relative number of failed students has shown that the failure rate has no linear trend. Therefore, the provided results in Figure 2 are only valid for children graduating with a grade of at least 4.0.

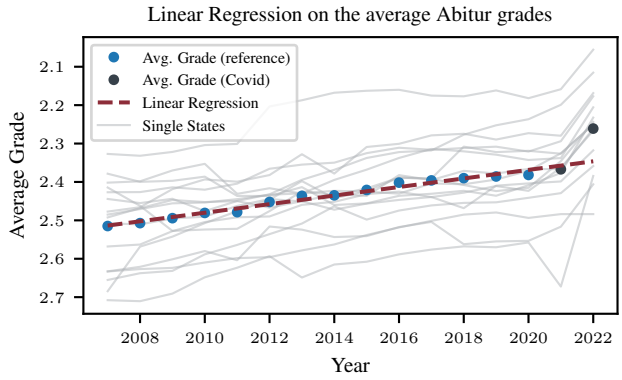

Figure 2. Average Abitur grades before (●) and after the COVID-19 pandemic (●) with a linear regression line (▬ ▬) of the years 2007 to 2020. In the background the figure contains average grades foreach federal state (_).

Correlation and Relationships Figure 3 presents a visualization of the Pearson correlation coefficients, analyzing the relationship between the number of children per teacher and the average number of repeaters, as well as the educational budget per state.

Initially, to compute the relative number of repeaters, the *Number of Repeaters* dataset is merged with the *school-children-by-state* dataset, grouped by *Federal State* and *Year* and the number of repeaters is divided by the total number of school children. Subsequently this dataset is merged with the *budget* and the *students by teacher* dataset, using *Federal States* and *Years* as the common attributes for alignment.

Finally, for each state, the Pearson correlation coefficient is calculated across different years to ascertain the correlation between the students-to-teacher ratio and the budget, as well as the average amount of repeaters.

In order to visualize the data over the states a heatmap for the german federal states is created. Therefor the Pearson correlation coefficients are normalized to the used colormap scale. Each state receives the appropriate computed color then.

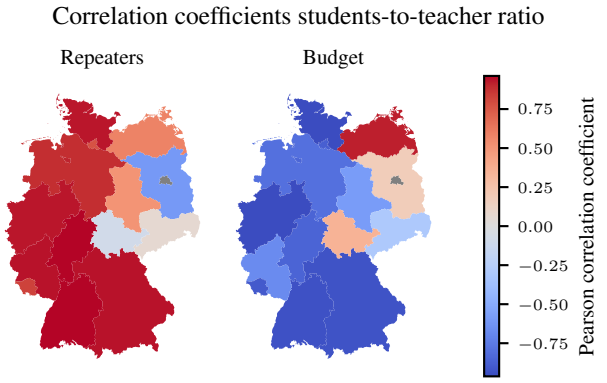

Figure 3. Pearson correlation coefficients between the student-to-teacher ratio and the relative repeater count (left) and the inflation-adjusted average budget per child (right). Red indicates positive, gray neutral, and blue negative correlations between the variables.

Results

The mostinteresting finding is the strong coreelation between the students per teacher ratio and the Abitur grades. Intuitevly this makes sense, since more teacher should result in smaller class sizes and less stress for the teachers, which makes for a better learning experience. The correlation between education performance and availabilty of teaching personnel is not new to research, but it was usually discussed in the context of university performance (Quelle). It is especially important, since the student teacher ratio got smaller over the current years and the grades went on a steep increase. This marks the importance of having enough teaching personnel for every school.

It would be easy to say that the schools just need more money, so they can employ more teachers, but that's not the whole story. In this case, thüringen, Sachsen-Anhalt and Brandenburg stand out. For them the correlation between budget and students per teacher is negative. This could be a classic case of the East-West-Gap in Germany. The exact reasons are up to speculation, but a plausible explanation might go like this: Schools get money mainly based on how many children they have -¿ If certain schools get more children, they might wanna employ more teachers -¿ Because they don't have enough teachers in these states, they can't -¿ Schools get more money, but the number of teachers stay the same: Negative correlation.

The same anomaly can be observed with the repeaters. Here, we think a different phenomenon is accountable for this. Schools in Thüringen and Sachsen-Anhalt rely more and more on Teilzeitkräfte. This means that the overall proportion of teachers increases, while the grades stay the same, or even worsen, becasue there is more flucuation in teaching presonell for a given class. Thus we get a negative

correlation.

That said, we can observe that for every other of the 16 federal states there is a very strong positive correlation, not only between Abitur grades but also the number of repeaters decreases. This means that it not only leads to better grades, but the weak ones won't left behind if enough teachers are available. But money doesn't necessaraly help here. There need to be enough teachers available to employ. From this analysis we conclude that making sure that many teachers are avilable is one of the most important challenges for the education system. As seen by the Studie Lehrerbedarfsdings bums by the german Kultusministerium... Was haben die eigentlich prognostiziert?

Conclusion

References

Grözinger, G. and Baillet, F. Gibt es auch beim abitur eine noteninflation? zur entwicklung der abiturnoten als hochschulzugangsberechtigung – eine darstellung und analyse aus soziologischer perspektive. *Bildung und Erziehung*, 68(4):473–494, 2015. doi: 10.7788/bue-2015-0407.

Jahnke, T., Klein, H. P., Kühnel, W., Sonar, T., and Spindler, M. Die hamburger abituraufgaben im fach mathematik. entwicklung von 2005 bis 2013. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 22(2):115–122, 2014. doi: doi:10.1515/dmvm-2014-0046. URL https://doi.org/10.1515/dmvm-2014-0046.

Kühnel, W. Modellierungskompetenz und problemlösekompetenz im hamburger zentralabitur zur mathematik. *Mathematische Semesterberichte*, 62:69–82, 2015.

Kultusminister Konferenz. Abiturnoten im Ländervergleich, 2024. URL https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html.

Lemmermeyer, F., Kühnel, W., Spindler, M., and Klein, H. P. Zentralabitur 2019: Weitere absenkung der mathematischen anforderungen zentralabitur 2019: The lowering of mathematical standards continous. *Journal für Didaktik der Naturwissenschaften und der Mathematik (F)*, 3: 92–98, 2019.

Rodgers, J. L. and Nicewander, W. A. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, 1988. ISSN 0003-1305. doi: 10.2307/2685263. URL https://www.jstor.org/stable/2685263. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Schleithoff, F. Noteninflation im deutschen schulsystem – macht das abitur hochschulreif? / grade inflation in the german school system. *ORDO*, 66(1):3–26, 2015. doi: doi:10.1515/ordo-2015-0103. URL https://doi.org/10.1515/ordo-2015-0103.

Statistische Bundesamt. Allgemeinbildende Schulen - Fachserie 11 Reihe 1 - Schuljahr 2020/2021 (Letzte Ausgabe - berichtsweise eingestellt), March 2022. URL https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/Publikationen/_publikationen-innen-schulen-allgemeinbildende.html.

Statistische Bundesamt. Statistisches Bundesamt Deutschland - GENESIS-Online, November 2023. URL https://www-genesis.destatis.de/genesis/online?language=de&sequenz=tabellen&selectionname=21111*#abreadcrumb.