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

This study analyzes the factors influencing the children's performance in the German school system through exploratory data analysis. Examining variables such as student numbers, teacher ratios, repeaters, and budget per child, the research aims to both describe and correlate these factors. The findings challenge the idea that grade inflation is the only cause of rising grades and emphasize the importance of having a small student-to-teacher ratio.

1. Introduction

The Abitur grades have constantly increased in the German school system over the past years (Kultusminister Konferenz), resulting in a research discussion in the media (todo source). The main topic of discussion is whether grades get better, even if students performance is worse.

The discourse has predominantly centered around mathematics, since the difficulty of exercises is easiest to compare. On the one hand-side, mathematicians argue that Germany has a grade inflation, resulting in easier exercises over the years (Kühnel, 2015; Jahnke et al., 2014; Lemmermeyer et al.). On the other hand-side, there are 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 by Grözinger & Baillet, involving the analysis of comprehensive data on the education system. The results were promising, but not yet enough to dismiss the claim of grade inflation.

This paper expands on that work, attempting to disprove the claim that grade inflation is the only cause of the observed trend. Building upon past research, a data analysis is conducted that aims to provide insight into the causes and effects on the schoolchildren's performance[1].

Furthermore, this study investigates whether there are any quantifiable causes of this upward trend in Abitur grades in the German education system. In addition, the causes and evolution of the repeater ratio across all school types are investigated to project the results to other educational institutions. Finally, prognoses about the trajectory of future grade developments and their implications for the German education system are outlined.

2. Methods

The basis of the exploratory data analysis are the used datasets and the applied mathematical concepts for forming a quantitative argumentation. Thus, this section first introduces the used data sets and explains afterward the mathematical background. Finally, this section analyzes first trends and findings on the effects of increasing grades.

2.1. Datasets

This paper investigates both the causes and effects of the phenomena discussed in section 1. Thus, the analyzed datasets are grouped into *cause* and *effect* datasets. All used datasets are collected on all public schools, since they are obliged to forward them to the federal statistical institutions (Statistische Bundesamt, 2024; Kultusminister Konferenz).

In the following, *causes* are defined as the social, demographic, or political factors that influence the German school system, such as the number of students, teachers, or budget provided by the German government. In contrast, *effect* refers to the observable impact of these causes on any measure modeling the students' performance, e.g., average grades or the rate of repeaters and school-leavers.

The first dataset is the *Fachreport Schuljahr 2020/21* presenting the number of teachers from 1992 until 2020 (Statistische Bundesamt, 2022). The dataset is grouped into contract type, federal state, and school type. In addition, the count

[*]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.

[1]Source files are publicly available in the project repository.

of schoolchildren is analyzed with two datasets (Statistische Bundesamt, 2024). The first dataset (21111−0002) contains the number of children per grade and school type for the years 1998 to 2022. In contrast, the second (21111 − 0010) provides the absolute number of children, leavers, and beginners in each federal state from 1997 to 2022. Therefore, the analysis can only be conducted separately for school types and federal states, due to the missing representation.

Additionally, this paper considers the budget per child (21711−0011) as a possible cause, which is provided by the Statistische Bundesamt. The dataset contains the budget per child for the years 2010 to 2022 and is grouped by federal states. To adjust for inflation, the budget is multiplied with the *Verbraucherpreisindex* (61111 − 0010) relative to 2022, as provided by the Statistische Bundesamt.

Furthermore, it is also important to analyze the effects on student performance, since they are the first indicator of whether grade inflation exists. One of the few publicly available datasets containing grades are the average Abitur grades per federal state. The grades are published every year in a separate reports by the Kultusminister Konferenz. Each file contains the count of children per written grade and federal state. The grades are given in increments of 0.1, with 4.0 being the worst and 1.0 being the best grade. The number 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 to translate the results. Accordingly, this paper uses the number of repeaters (21111 − 0014) from the Statistische Bundesamt. There, the absolute count of repeaters by federal state, school type, and year is provided for the years 1998 to 2022.

2.2. Mathematical Concepts

The subsequent paragraph provides an overview of the two primary used concepts of this paper: linear regression and the Pearson correlation coefficient.

Firstly, linear regression is a statistical technique for modeling the relationship between a dependent variable and one or more independent variables (James et al., 2021). It seeks to fit a linear equation to the data that minimizes the discrepancy between observed and predicted values. This is done through least squares minimization, resulting in an equation of the form (James et al., 2021):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon \qquad (1)$$

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

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

2.3. Exploratory Data 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 of the teachers and children datasets[2] 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.

Moreover, this demographic effect is combined with an increasing number of teachers across all German school types and federal states. Although, there is a difference between old and new federal states in the evolution of full- and part-time teachers, as displayed in Figure 1. The new federal states had a higher proportion of part-time teachers than full-time teachers between 2004 and 2008. Instead, the old federal states have a nearly constant distribution of part-time teachers on average. Furthermore, they have employed fewer full-time teachers since 2014 than the new federal states.

Given the hypothesis that having more teachers per student increases the quality of teaching, the datasets are merged (Kasau Onesmus Mulei et al., 2016; Koc & Celik, 2015). 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 ($\sim 90\%$) of the distribution (Figure 1). As a result, the average over the 5 most common schools 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 analysis of the adjusted budget[3] 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 of schools in the last years (Cone et al., 2022).

Now that some basic effects that may influence the students'

[2]Teachers and children plots and exploration can be found here.
[3]Budget plots and exploration can be found here.

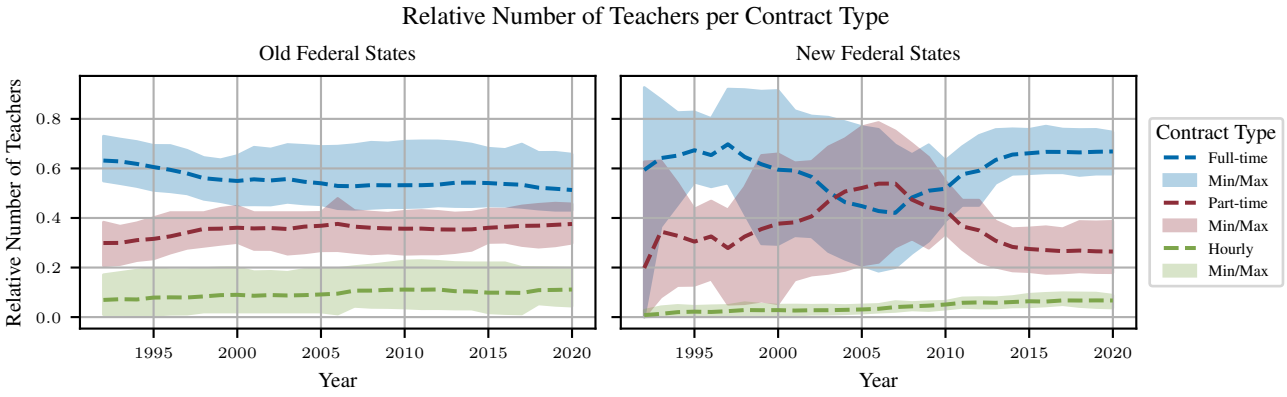

Figure 1. Relative number of teachers by contract type and federal states. The average ratio of full-time (▬ ▬), part-time (▬ ▬) and hourly-based (▬ ▬) contracts of teachers is displayed with its minimal and maximal bound for the old federal states (*left*) and the new federal states (*right*).

performance have been identified, it is possible to study the performance measures. As the analysis of the students datasets[2] 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. The analysis results of the Abitur grades[4] are shown in Figure 2. 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[4] of the relative number of failed students has shown that the failure rate has no repetitive or linear pattern. Therefore, the provided results in Figure 2 are only valid for children graduating with a grade of at least 4.0.

Moreover, all children attending other schools have no direct impact on the results of the Abitur grades. Therefore, the number of repeaters per federal state, school type, grade, and school year is analyzed. To enhance the relevance of the results, the relative ratio of repeaters[5] is calculated by dividing the absolute counts by the absolute number of schoolchildren. This results in an aggregation for the federal states per year and in one for the school types per year. As a result, the number of repeaters has decreased for all educational institutions and federal states from 1998 to 2020. Hence, the trend equals the expected result, after analyzing the Abitur grades.

To summarize, the exploratory findings indicate an increasing number of students and teachers, resulting in a decreasing ratio of students to teachers and a rise in the budget per child. The possible outcomes include a linear increase in Abitur grades in grammar schools and a shrinking proportion of repeaters in general.

[4]Abitur plots and exploration can be found here.
[5]Repeater plots and exploration can be found here.

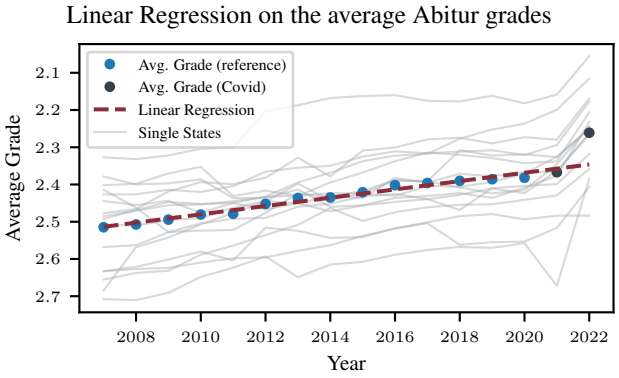

Figure 2. Average Abitur grades before (●) and after the COVID-19 pandemic (●) with a linear regression line (▬ ▬) of the years 2007 to 2020. In the background the figure contains average grades for each federal state (▬).

3. Results and Discussion

The introduced student-to-teacher ratio combines the number of children with the number of teachers, either for each federal state or school type. Thus, this section compares the student-to-teacher ratio with the Abitur grades, repeaters, and budgets. Importantly, the Abitur grades can only be analyzed for the German average, due to the missing representation. Nevertheless, the repeaters and budgets are analyzed for each federal state.

One of the key findings of this analysis is the strong correlation between the average grades across all federal states and the student-to-teacher ratio in German grammar schools. As shown in Figure 3, the relationship between both is nearly

linear and contains neither clusters nor outliers. Hence, a smaller student-to-teacher ratio strongly correlates to better Abitur grades, with a Pearson correlation coefficient of 0.98.

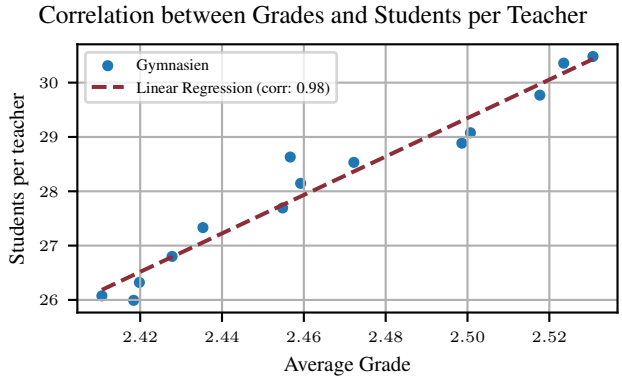

Figure 3. Linear regression on the student-to-teacher ratio by average Abitur grade. The resulting regression line (▬) is calculated over the aggregated average overall grammar schools (●) in Germany.

This result confirms the initial hypothesis of the big influence of the student-to-teacher ratio and is consistent with studies in other countries (Kasau Onesmus Mulei et al., 2016; Koc & Celik, 2015; Dickson, 1984).

However, the Abitur grades give only an insight into grammar schools. Thus, the correlations with repeaters and budgets are shown in Figure 4. It 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 child. Therefore, the Pearson correlation coefficients for each state are normalized to the used color map scale.

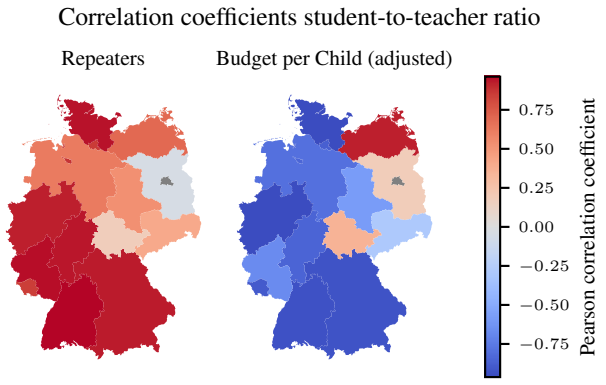

Figure 4. Pearson correlation coefficients between the student-to-teacher ratio and the relative repeater count (left) and the inflation-adjusted average budget per child (right). Red indicates positive, light gray neutral, gray missing, and blue negative correlations between the variables.

The findings represented in Figure 4 support a positive correlation between the student-to-teacher ratio and the number of repeaters in most federal states. In contrast, they indicate a negative correlation between the student-to-teacher ratio and budget per child for the same states. Furthermore, there is a big difference between *old* and *new* federal states for both correlations. Especially, the results for Thüringen and Brandenburg diverge from the average in both maps. Rather, Mecklenburg-Vorpommern differs most in the budget per child from the other states.

However, an increase in the number of students in the new federal states can explain the different correlations in the budget per child[2]. Since the per-child budget has increased in all states[3], the schools got more cash in total. This results in more vacancies at schools (Kultusminister Konferenz, 2023) and other investments, like digitalization and maintenance of schools (Bundesministerium für Bildung und Forschung, 2022). But in the short term, there will be more teachers needed as available (Kultusminister Konferenz, 2023). As Figure 1 indicates, this leads to a high variance in the distribution of teacher contract types, especially for the new federal states. This results in a higher student-to-teacher ratio with the same budget.

In contrast, the anomaly for the repeaters involves more aspects of the educational system, like the different curricula or conditions for repeating a class. The importance attached to repetition varies from federal state to federal state and cannot necessarily be used as an indicator of educational quality (Klemm, 2009). Hence, a complete explanation for these results involves more datasets and effects, which are not analyzed in this paper.

Moreover, all other states indicate that a higher budget per child results in a smaller student-to-teacher ratio. As a result, the number of repeaters decreases with the student-to-teacher ratio.

4. Conclusion

In summary, this exploratory data analysis found correlations between the ratio of children and teachers, Abitur grades, repeaters, and budget per child. These indicate the importance of employing enough teachers, to increase children's performance. Prognoses are showing that there are still more open positions than teachers that can fill them (Kultusminister Konferenz, 2023). Unfortunately, they predict that this gap will eventually close in the coming decade.

Furthermore, it is evident from the observed correlations that the grades should get worse in the short term and increase in the long term, if there are no other factors influencing the Abitur grades. All in all, this analysis shows that grades at schools do not only rise due to grade inflation, but are also influenced by other positive factors.

References

Bundesministerium für Bildung und Forschung. Fortschrittsbericht DigitalPakt Schule 2019–2022, June 2022. URL https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31715_Fortschrittsbericht_DigitalPakt_Schule_2019_bis_2022.pdf?__blob=publicationFile&v=6.

Cone, L., Brøgger, K., Berghmans, M., Decuypere, M., Förschler, A., Grimaldi, E., Hartong, S., Hillman, T., Ideland, M., Landri, P., Van De Oudeweetering, K., Player-Koro, C., Bergviken Rensfeldt, A., Rönnberg, L., Taglietti, D., and Vanermen, L. Pandemic Acceleration: Covid-19 and the emergency digitalization of European education. *European Educational Research Journal*, 21(5): 845–868, September 2022. ISSN 1474-9041, 1474-9041. doi: 10.1177/14749041211041793.

Dickson, V. A. An Economic Model of Faculty Grading Practices. *The Journal of Economic Education*, June 1984. Publisher: Taylor & Francis.

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, July 2015. doi: 10.7788/bue-2015-0407. Publisher: Böhlau Verlag.

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), January 2014. ISSN 0942-5977, 0947-4471. doi: 10.1515/dmvm-2014-0046.

James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer US, New York, NY, 2021. ISBN 978-1-07-161417-4 978-1-07-161418-1. doi: 10.1007/978-1-0716-1418-1.

Kasau Onesmus Mulei, Kaloki Joseph Waita, Kitoo Beth Mueni, Mutinda Julius Mutune, and Dr. Jeremiah Kalai. Pupil-Teacher Ratio And Its Impact On Academic Performance In Public Primary Schools In Central Division, Machakos County, Kenya. June 2016. doi: 10.5281/ZENODO.54767. Publisher: Zenodo.

Klemm, K. Klassenwiederholungen teuer und unwirksam, 2009. URL https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/klassenwiederholungen-teuer-und-unwirksam.

Koc, N. and Celik, B. The Impact of Number of Students per Teacher on Student Achievement. *Procedia - Social and Behavioral Sciences*, 177:65–70, April 2015. ISSN 1877-0428. doi: 10.1016/j.sbspro.2015.02.335. URL https://www.sciencedirect.com/science/article/pii/S1877042815016894.

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

Kultusminister Konferenz. Lehrkräfteeinstellungsbedarf und -angebot in der Bundesrepublik Deutschland 2023-2025, August 2023.

Kühnel, W. Modellierungskompetenz und Problemlösekompetenz im Hamburger Zentralabitur zur Mathematik. *Mathematische Semesterberichte*, 62(1):69–82, April 2015. ISSN 0720-728X, 1432-1815. doi: 10.1007/s00591-015-0145-9.

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.

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. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Schleithoff, F. Noteninflation im deutschen Schulsystem - Macht das Abitur hochschulreif? In *ORDO*, pp. 3–26. De Gruyter, December 2015. ISBN 978-3-11-050604-4. doi: 10.1515/9783110506044-002.

Statistische Bundesamt. Allgemeinbildende Schulen: Schuljahr 2020/2021. volume 1 of *Fachserie 11*. Statistisches Bundesamt, 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, January 2024. URL https://www-genesis.destatis.de/genesis/online.