Grading in Secondary Schools in Germany - The Impact of Social Origin and Gender

Table S1. Means, shares, standard deviation shares of nonresponse of all variables in the multivariate model

variables	mean resp. shares[e] (unweighted)	standard deviation[e] (unweighted)	shares of nonresponse[e] (unweighted)	mean resp. shares[e] (weighted)	standard deviation[e] (weighted)	shares of nonresponse[e] (weighted)
Level: students[a]						
Competency score in German orthography	-0,06	1,14	22,98 %	-0,08	1,11	21,65%
BEFKI (general cognitive abilities)	-0,06	1,17	1,02 %	-0,08	1,17	0,87%
Effort level						
Item: *Hours on a problem because I cannot rest without answer*	3,23	1,12	54,37 %	3,23	1,15	54,51%
Item: *Finish things that I begin with*	3,56	1,01	54,51 %	3,59	1,02	54,63%
Item: *If I can not solve a problem, I work even harder*	3,40	0,99	54,46 %	3,43	1,00	54,63%
Item: *Work like a fiend at problems that I feel must be solved*	3,42	1,04	54,61 %	3,42	1,05	54,65%
Gender						
female	0,50	--	0,00 %	0,49	--	0,00%
male	0,50	--	0,00 %	0,51	--	0,00%
Student has special educational needs						
yes	0,03	--	0,06 %	0,01	--	0,01%
no	0,97	--	0,06 %	0,99		0,01%
Mother with tertiary educational qualification						
yes, university	0,14	--	20,58 %	0,12	--	19,96%
yes, college resp. bachelor	0,09	--	20,58 %	0,08	--	19,96%
no	0,77	--	20,58 %	0,80	--	19,96%
Student has migration background						
yes	0,25	--	8,79 %	0,29	--	8,74%
no	0,75	--	8,79 %	0,71	--	8,74%
Interaction: female and mother with university degree						
yes	0,07	--	20,58 %	0,06	--	19,96%
no	0,93	--	20,58 %	0,94	--	19,96%

Interaction: female and mother with college resp. bachelor degree						
yes	0,04	--	20,58 %	0,04	--	19,96%
no	0,96	--	20,58 %	0,96	--	19,96%
Interaction: competency and mother with university degree						
yes	0,07	0,44	38,80 %	0,06	0,40	37,53%
no	0,02	0,96	38,80 %	-0,01	0,96	37,53%
Interaction: competency and mother with college resp. bachelor degree						
yes	0,02	0,32	38,80 %	0,03	0,99	37,53%
no	0,07	1,01	38,80 %	0,03	0,99	37,53%
Level: school resp. class (b)						
Share of students with special educational needs in school	16,75	23,63	13,26 % (f)	12,32	18,30	14,75% (f)
Share of students in class	22,17	4,80	0,00 %	22,24	4,96	0,00%
Share of students with migration background in class	0,50	0,41	0,28 % (g)	0,56	0,40	0,21% (g)
Share of female students in class	0,49	--	0,00 %	0,48	--	0,00 %
Mean level of competency in class	-0,15	0,82	16,42 %	-0,17	0,78	15,99 %
Class in *Gymnasium* school type						
yes	0,61	--	0,00 %	0,45	--	0,00%
no	0,39	--	0,00 %	0,55	--	0,00%
School structure of federal states						
Modernized structure	0,63	--	0,00 %	0,49	--	0,00%
Mixed modernized / mixed traditonal	0,37	--	0,00 %	0,51	--	0,00%
Level: teacher (c)						
Professional experience (years)	17,30	12,47	0,00 %	14,74	11,44	0,00%
Teacher gender						
female	0,75	--	0,00 %	0,71	--	0,00%
male	0,25	--	0,00 %	0,29	--	0,00%
Lateral entry						
yes	0,12	--	0,00 %	0,14	--	0,00%

no	0,87	--	0,00 %	0,84	--	0,00%
Level: students and teacher						
Interaction: teacher is female and student is female (d)						
yes	0,38	--	0,00 %	0,35	--	0,00%
no	0,62	--	0,00 %	0,65	--	0,00 %

Notes: (a) the descriptive values for students refer to the sample of N = 31,594 students whose missing values were imputed for the analyses. For the weighted analyses we use the original sample weights out of the IQB Trends in Student Achievement study (*totwgt_deu*). (b) The values for class and school characteristics refer to the related class and school sample of size N = 1,425. Here we are using also the original sample weights for the analyses. (c) Values on the level of teachers are referring to the reduced sample of N = 21,813 students, where N = 978 teachers could be assigned. For the other students it was impossible to assign the referring teacher. In most of these cases the teacher did not participate. In some cases, it is possible that the information for connecting student and teachers has lost, but we could not distinguish between these cases. For teachers we are using our own adjusted weights on all imputed datasets and generating the average of all calculated means, standard deviations resp. shares. (d) The descriptive values of this interaction effect is calculated for N = 21,813 students, where teacher information is available. (e) In calculating the unweighted means, shares, and standard deviations we ignore missing values. (f) For 189 schools we have no information about this. (g) There is no information about ethnic background for students in these classes.

S2. Methodological Notes

Multiple imputation of missing student attributes

All missing values in the student data (with $N = 31{,}594$ cases) were replaced using multiple imputation with *chained equations* (van Buuren and Groothuis-Oudshoorn 2010). Several reasons for proceeding in this way exist: restricting the analysis to complete data records only would reduce the explanatory potential of the data considerably. Analyzing complete data records only would also imply that it can be assumed that the missing data is missing completely at random (MCAR). This is clearly unrealistic; we need only think of the information on mothers' educational backgrounds that is only available when it has been supplied via the parent questionnaire. Thirdly, we can compute class aggregates for incompletely observed variables (such as migration background) proceeding in this way. As participation in the IQB Trends 2015 is obligatory for all students, at least partial information is available for almost all students in the sampled classes. The risk of students being missing (through illness, for example) is low, random, and can be regarded as essentially negligible. We used the two-level predictive mean matching imputation method implemented in the R package *miceadds* (Version 3.5-14) to impute missing values in the student data (Robitzsch et al. 2019). This method allows adequate consideration of variable types and of the hierarchical structure of the data, which is generated automatically by siting the students within classes and schools. Neglecting this specific data structure could distort the estimators of the parameters of interest (Drechsler 2015). Following a recommendation by von Hippel (2009), we additionally replaced the missing values in the interactions between the variables *educational background of mother* and *student gender* and between *educational background of mother* and *student competency level* by imputing missing data (in the relevant imputation step) only after stratifying the data by gender or by education categories. We replaced missing values relating to school attributes using a one-level imputation procedure. We used considerably more variables to predict missing values in our imputation models than we considered in our analysis model. Proceeding in this way serves to enhance the robustness of the predictions of the imputation models as far as possible (see Collins et al. 2001).[i] In total, 20 datasets were imputed.

Nonresponse adjustments for missing teacher information

We adjusted the survey weights (provided with the 2015 IQB Trends data) in each of these imputed datasets to compensate for information missing at the teacher level. Specifically, we formed aggregates at class level from the student attributes for every imputed dataset and drew on these in addition to the information available for schools (like the *education system variable* constructed beforehand*, school size*, *population of school location* and *proportion of students at school with German as a native language*) to fit adjusted models with regard to the availability of teacher information for a student. The results of these models were then pooled using Rubin's Rules (Rubin 1987) to reach an overall result (see Table S2). It is clear from the results (despite the uncertainty introduced by imputing missing values in the adjustment models) that the variables and auxiliary variables *proportion of girls in class*, *average English reading competency in class, mean general cognitive abilities of class*, *class is a Gymnasium class, education system of federal state*, *school provides all-day schooling* and *population of school location* have significant and substantial effects on the availability of teacher information. It is thus clear that the missing teacher information is not missing completely at random (MCAR). Ignoring this fact could distort the regression weights in the estimation of the analysis model (Little and Rubin 2019). We thus used suitable adjustment weights to compensate for the non-MCAR pattern of missing teacher information. To achieve this, we estimated the probability of teacher information being available for every student in every imputed dataset. At this, we only used the regressors that had proved to be significant in the overall results of the missing data analysis for this purpose. (We counteracted variance in the weights between the imputed datasets in this way). The inverse values of these estimated probabilities represent the adjustment factor for correcting the sampling weights in the IQB Trends in Student Achievement 2015 data to compensate for missing teacher information. Using inverse probabilities to compensate for unit nonresponse in data is termed propensity score weighting (Rosenbaum and Rubin 1983) and it is a common technique in survey statistics (Little and Vartivarian 2005).

Fitting the analysis model

We were then able to derive aggregates for the class context (such as the proportion of students with migration backgrounds) from the student weights derived in this way and to additionally incorporate these, along with the adjusted weights, into our analysis model (i.e. the weighted *ordered logit* model). The results of the weighted analyses, which we computed for the imputed and weighted datasets in each case, were pooled using Rubin's Rules.[ii] The strategy of combining multiple imputation and propensity score weighting is not entirely new and has already been applied in similar data situations, for example by Seaman et al. (2012) and Quartagno et al. (2019).

To ensure that our results can be replicated, the entire data preparation, imputation and weighting syntax in R and the syntax for fitting the models in Stata and combining the results using Rubin's Rules is publicly available and may be downloaded from … <Link will be added in the final manuscript>.

Table S2. Estimated (non-standardized) beta-coefficients with 95% confidence intervals of the nonresponse model of teacher variables

variables	reference	effect (non-standardized)	95 % confidence interval
Share of students in class	--	0,008	(0,000; 0,016)
Share of female students in class	--	0,468*	(0,254; 0,683)
Share of mothers with tertiary educational qualification	--	-0,048	(-0,716; 0,619)
Share of fathers with tertiary educational qualification	--	-0,323	(-0,996; 0,351)
Mean age in class		-0,041*	(-0.057; -0.025)
Mean grade in German in class	--	0,003	(-0,154; 159)
Mean cognitive abilities (BEFKI) in class	--	-0,354*	(-0,485; -0.222)
Mean competency in class in	--		
reading		0,120	(-0,127; 0,366)
listening comprehension		0,124	(-0,182; 0,430)
orthography		0,112	(-0,063; 0,286)
reading in English		0,280*	(0,168; 0,391)
listening comprehension in Englisch		-0,089*	(-0,154; -0,024)
Class in *Gymnasium* school type	no		
yes		1,185*	(1,063; 1,307)
Number of students in ninth grade		0,000	(-0,002; 0,001)
Number of students with special educational needs in school	--	0,001	(-0.002, 0,003)
Share of students with German as first language in school	less than 25%		
26 to 50%		0,051	(-0,265; 0,367)
51 to 75%		-0,050	(-0,308; 0,208)
76 to 90%		0,104	(-0,166; 0,374)
More than 90%		0,143	(-0,130; 0,415)
Educational system of federal state	Mixed modernized / mixed traditional		
modernized		-0,377*	(-0,447; -0.307)
Population of place of school	less than 3000		
until 15000		0,035	(-0,180; 0,250)
until 50000		-0,214	(-0,435; 0,027)
until 100000		-0,176	(-0,445; 0,092)
until 500000		-0,318*	(-0,551; -0,085)
more than 500000		-0,224	(-0,475; 0,027)
School is full-time school	no		
yes		-0,191*	(-0,308; -0,075)
Sample size			
Level of students	31.594		
Level of class/school	1425		
R^2 (Nagelkerke)	0,155		

Comment: Estimated model: logistic regression with *glm* function of the R version 3.5.1. Dependent variable: 1 information from teacher / 0 no information from teacher. * significant on $p<0.05$.

References

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological methods, 6(4), 330.

Drechsler, J. (2015). Multiple imputation of multilevel missing data—Rigor versus simplicity. Journal of Educational and Behavioral Statistics, 40(1), 69–95.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? Survey Methodology, 31(2), 161–168.

Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.

Quartagno, M., Goldstein, H., & Carpenter, J. R. (2019). Multiple Imputation with survey weights: a multilevel approach. Journal of Survey Statistics and Methodology, 8(5), 965–989.

Robitzsch, Alexander, Grund, Simon, Henke, Thorsten (2019). miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'. R package version 3.5-14. https://CRAN.R-project.org/package=miceadds.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. Hoboken: John Wiley & Sons.

Seaman, S. R., White, I. R., Copas, A. J., & Li, L. (2012). Combining multiple imputation and inverse‐probability weighting. Biometrics, 68(1), 129‐137.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011) mice: multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1–67. https:// doi.org/10.18637/jss.v045.i03

von Hippel, P. T. (2009). 8. How to impute interactions, squares, and other transformed variables. Sociological methodology, 39(1), 265–291.

---

[i] Specifically, we used the following variables at school or class level in the imputation models: the population of the school location, the proportion of students at the school with German as their native language, the total number of ninth-grade students in the school, whether the school provided all-day schooling in the school year 2014/15, the proportion of students with special educational needs at the school, the education system of the federal state in the form of the education system categories constructed beforehand, the mean competency of the class in the two competency domains studied and the mean general cognitive abilities of the class. At the student level we used: the half-year grade in German, the educational level of the mother and the father, the student's migration background, the special needs diagnosis (if any existed), the score for the general cognitive abilities, the various competency scores available in the 2015 IQB Trends data, the four items on intellectual curiosity (that we used to construct the factor for

a student's effort level), and age. Apart from that, we used all analysis variables of the analysis model and their aggregates at the class level for imputing missing values as well.

[ii] As an alternative to this kind of modeling, the variables showing significant and strong effects could theoretically also be integrated into the analysis model (a model-based approach) to avoid bias in estimating the model. However, the model we are looking at contains many analysis variables at multiple levels. Adding further (control) variables would lead to convergence issues in estimating the model due to low cell counts. To avoid this problem, we have instead opted to use the design-based approach described above to compensate for missing teacher information.