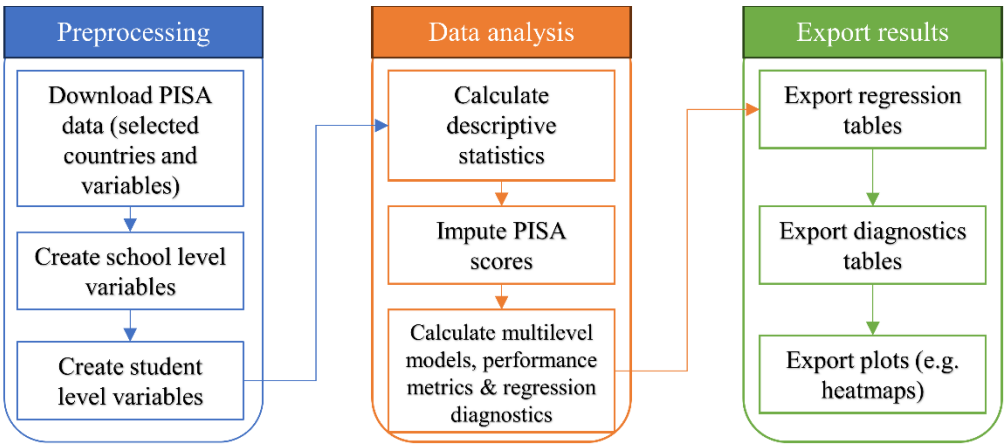
Online Appendix

This online appendix of the monograph entitled "Effektive Schulsteuerung? Bilanz einer globalen Reformagenda" is divided into four parts. In part A, we provide a definition of variables included in the regression analysis and the robustness checks. In part B, we continue with descriptive statistics. Part C contains the full regression tables referred to in the monograph. In part D, we introduce regression diagnostics and metrics, namely the Variance Inflation Factor (VIF) and the Shapiro-Francia W' Test. For data preparation and calculation procedures, we used the R version 4.3.2, RStudio version 2024.4.2764, and the packages listed in table 1 (with version number and purpose of application). The workflow is depicted in figure 1 and provides an overview of the steps taken to preprocess and analyze the OECD PISA data.

Table 1: Overview of the R packages, including version number and intended purpose, which were used for data preparation and calculations.

R package	Version number	Intended Purpose
haven	2.5.3	Read the OECD SPSS-files for the PISA 2000 wave
EdSurvey	4.0.4	Download the OECD data for PISA waves 2009 and 2015
xlsx	0.6.5	Package used to read and write excel files
tidyverse	2.0.0	Data preprocessing and data frame construction
dplyr	1.1.3	
stringr	1.5.0	
purr	1.0.2	
reshape2	1.4.4	Convert wide format of PISA-data frame to long format and vice versa
psych	2.3.9	Used to generate a wide range of descriptive statistics
BIFIEsurvey	3.5.19	Employed to impose the OECD weighting scheme on the data and to impute the PISA scores in math, reading, science, and the overall score
lme4	1.1.34	Used to calculate multilevel linear regression models
nortest	1.0.4	Used to test for the violation of the normality of the error terms assumption
performance	0.10.8	Used to calculate Snijders and Bosker's R^2 and ICC-values for all multilevel linear regression analyses
stargazer	5.2.3	Generate a table output of descriptive statistics and regression models
ggplot2	3.4.4	Used to design and visualize graphics based on the calculations
dotwhisker	0.8.2	Used to visualize an effect plot of the main effects provided by the multilevel linear regression models
effect	4.2.2	

Figure 1: Workflow employed to preprocess and analyze the OECD PISA data.



A Variables Included in the Regression Analysis

We used the imputed mean PISA score as dependent variable to avoid bias in the estimation of the association between independent variables and the PISA scores. The calculation of the mean PISA score was as follows. In each wave under scrutiny (2000, 2009, 2015), the PISA test was split into different question clusters (e. g. S1 to S5 for science, M1 to M5 for mathematics and R1 to R5 for reading in 2000). Not every question cluster was administered to every student. Instead, multiple imputation methods were applied to estimate whether a student might have given correct answers in unadministered question clusters. The basis for the imputation was a latent regression model with social context variables as predictors (OECD, 2017, pp. 145–149 and 181–182). Following this step, a principal component analysis was conducted to remove collinearity among variables where present. Finally, weighting schemes on school and student level were applied to provide means to calculate the standard errors of estimators correctly and to calculate sampling variance (OECD, 2017, pp. 116-126). To do so, the OECD applied a balanced repeated replication with Fay's method (Judkins 1990) and provided 80 separate data matrices and grouped schools with similar characteristics. Based on this calculation, five plausible values for the science, reading, and math scores were predicted in 2000 and 2009, and 10 in 2015. Using the R-package BIFIEsurvey (Robitsch and Oberwimmer, 2019), we first calculated the mean plausible values for each subject and then imputed the scores to attain the final mean PISA scores which were then used as dependent variable.

For measuring the socioeconomic background of students and schools, we used the Index of Economic, Social, and Cultural Status (ESCS). The ESCS is a measure designed to capture the socioeconomic status of a household and relates it to measured student performance in the PISA test. According to Avvisati (2020, p. 3), it is defined as follows:

"ESCS is a measure of students' access to family resources (financial capital, social capital, cultural capital and human capital) which determine the social position of the student's family/household."

ESCS is a composite variable generated by a factor analysis on the parents' highest occupational status, a parent's highest level of education, and home possessions reported by the students. These are possessions that indicate material well-being (e.g., availability of one's own room), cultural affinity (availability of a musical instrument, classical literature, paintings), and educationally relevant resources (e.g., a quiet place to study). In addition, the number of available books in the household is also included in the calculation of the ESCS. Mathematically, this is the summation of the parents' highest occupational status, their highest educational attainment, and home possessions which are divided by the eigenvalue of the first factor generated by the factor analysis (OECD, 2017, pp. 339–340). In the process, the possessions captured by the home possession items changed slightly. For example, in later waves, students were asked about the availability of their own computer for school work or for general use, and other variables were added from 2009 onward (OECD, 2012). The ESCS variables were weighted based on the previous waves, so that at least a rough comparability is given, and they must be interpreted with due caution.

The school-level ESCS was constructed by calculating the mean of the student-level ESCS. The calculation was based on all students sampled for the respective school.

We furthermore included migration background (second generation) and language spoken at home to address the association between a family's migration background and the PISA score.

The indices of leadership, autonomy, and accountability, included in the models, were calculated as follows for the respective years:

2000:

Leadership is constructed as the proportion of the answer "principal" when asked about responsibility for 12 decision-making competencies.

Autonomy is defined as the ratio of mentions of "principal" and "teachers" to "regional or local education authority" and "national education authority" for items on responsibility for the following six decision-making competencies: hiring teachers, firing teachers, setting entry-level teacher salaries, increasing teacher salaries, formulating school budgets, allocating school budgets.

Accountability is constructed as additive index of "yes" responses to the six questions on control functions (ranged 0–6, listwise deletion of missing values).

2009:

The leadership scale "is based on fourteen items. Positive WLE scores on this index indicate greater involvement of school leadership in school affairs." (OECD, 2012, Items SC26Q01–SC26Q14).

Autonomy is defined as the ratio of mentions of "principal" and "teachers" to "regional or local education authority" and "national education authority" for items on responsibility for the following six decision-making competencies: hiring teachers, firing teachers, setting entry-level teacher salaries, increasing teacher salaries, formulating school budgets, allocating school budgets.

Accountability is constructed as additive index of "yes" responses to the eight questions on the use of tests for control functions (ranged 0–8, listwise deletion of missing values).

2015:

Leadership is constructed using the SC009 item battery which included 13 items on school leadership. According to the PISA technical report 2015, "The results provided data for five scaled indices. Principals were asked to indicate the frequency of the listed activities and behaviours in their school during the last academic year. The six response categories were 'did not occur', '1-2 times during the year', '3-4 times during the year', 'once a month', 'once a week', to 'more than once a week'. The overall scale for leadership (LEAD) consists of all 13 items. Table 16.53 shows the item wording, international item parameters and item fit for LEAD." (OECD, 2017, p. 324)

To calculate school autonomy, the PISA index of school autonomy (SCHAUT) is applied: „The index of school autonomy (SCHAUT) is calculated as the percentage of tasks included in SC010 (yes/no questions) for which the principal, the teachers or the school governing board have considerable responsibility." (OECD, 2016, p. 243).

Accountability is constructed as additive index of disjunctions containing "yes" responses of items of use of standardized or individual teacher-designed tests for various control functions (ranged 0–6, listwise deletion of missing values).

B Descriptive Statistics

We continue with the descriptive statistics of the variables included in the regression models and provide the point estimates and dispersion of the variables in tables B1-B12. Tables B1-3 refer to the UK, B4-6 to Germany, B7-9 to Sweden, and B10-12 to Finland. As point estimates, we calculated the minimum, mean, maximum, and median. Skewness and kurtosis were calculated to depict the dispersion of the dependent and independent variables included in our regression models. We furthermore included heatmaps depicting the correlation among each variable included in the model or robustness checks in figures B1-B12. Like the tables, figures B1-3 refer to the UK, B4-6 to Germany, B7-9 to Sweden, and B10-12 to Finland. Note that the data for UK (tables and figures 1-3) does not include Scotland.

Table B1: Descriptive statistics of the independent and dependent variables used for the regression models for UK in 2000. Own calculations.

Summary UK 2000	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	3850	529.783	100.211	531.77	152.19	824.66	-0.161	-0.116	1.615
Plausible value 1: reading	6969	521.908	102.374	525.98	97.54	846.89	-0.216	-0.048	1.226
Plausible value 1: mathematics	3878	526.893	93.64	531.58	128.93	819.15	-0.26	0.123	1.504
Plausible value 2: science	3850	528.857	100.404	533.945	137.76	838.66	-0.159	-0.257	1.618
Plausible value 2: reading	6969	522.156	102.597	524.88	137.54	881.96	-0.212	-0.087	1.229
Plausible value 2: mathematics	3878	527.091	93.755	529.12	169.38	837.47	-0.247	-0.001	1.506
Plausible value 3: science	3850	529.658	99.103	532.71	151.72	841.22	-0.157	-0.18	1.597
Plausible value 3: reading	6969	521.656	102.049	526.56	144.81	826.81	-0.235	-0.106	1.222
Plausible value 3: mathematics	3878	526.902	93.768	531.19	166.36	796.59	-0.252	-0.008	1.506
Plausible value 4: science	3850	529.063	99.927	531.79	160.74	814.12	-0.151	-0.194	1.61
Plausible value 4: reading	6969	521.646	102.025	526.7	136.05	826.51	-0.216	-0.065	1.222
Plausible value 4: mathematics	3878	526.72	93.992	531.27	184.65	802.87	-0.248	-0.063	1.509
Plausible value 5: science	3850	529.566	100.534	533.3	180.52	833.06	-0.176	-0.251	1.62
Plausible value 5: reading	6969	521.831	102.223	525.98	127.87	835.07	-0.237	-0.078	1.225
Plausible value 5: mathematics	3878	527.376	93.935	532.51	141.26	796.11	-0.261	0.033	1.508
Student level: ESCS	6797	0.223	0.83	0.216	-2.68	2.902	-0.062	-0.422	0.01
School level: ESCS	6969	0.216	0.435	0.205	-0.978	1.43	0.099	-0.109	0.005
Student level: Migrant background second generation	6969	0.158	0.364	0	0	1	1.879	1.533	0.004
School level: % Migrant background second generation	6969	0.158	0.136	0.125	0	0.828	2.156	6.182	0.002
Language at home = Test language	6838	0.972	0.164	1	0	1	-5.745	31.008	0.002
Student level: School disciplinary climate (WLE)	6969	1.217	10.88	-0.18	-2.92	97	8.61	72.839	0.13
School level: Mean disciplinary climate	6969	1.217	3.385	0.079	-1.325	29.017	4.38	24.134	0.041
School autonomy	6363	1.129	0.637	0.98	-0.29	1.72	-0.537	-1.109	0.008
Educational leadership	6221	0.001	0.011	0	0	0.1	8.94	77.933	0
Accountability	6172	5.045	0.906	5	1	6	-0.699	-0.161	0.012

Table B2: Descriptive statistics of the independent and dependent variables used for the regression models for UK in 2009. Own calculations.

Summary UK 2009	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	9548	508.971	98.559	511.69	113.8	816.43	-0.124	-0.176	1.009
Plausible value 1: reading	9548	490.722	95.041	493.16	115.01	812.02	-0.164	-0.113	0.973
Plausible value 1: mathematics	9548	487.672	85.969	488.4	133.44	807.61	-0.079	-0.095	0.88
Plausible value 2: science	9548	509.219	98.41	511.41	91.89	842.73	-0.111	-0.094	1.007
Plausible value 2: reading	9548	491.077	95.019	494.26	78.47	788.98	-0.145	-0.089	0.972
Plausible value 2: mathematics	9548	487.345	86.029	487.47	107.66	802.47	-0.066	-0.048	0.88
Plausible value 3: science	9548	509.205	98.98	510.57	100.75	884.69	-0.098	-0.107	1.013
Plausible value 3: reading	9548	490.888	95.613	492.63	109.56	799.48	-0.138	-0.067	0.978
Plausible value 3: mathematics	9548	486.936	86.98	486.3	78.68	785.02	-0.069	-0.073	0.89
Plausible value 4: science	9548	509.008	99.145	510.76	135.9	848.32	-0.104	-0.155	1.015
Plausible value 4: reading	9548	490.921	95.298	493.18	132.01	789.85	-0.15	-0.147	0.975
Plausible value 4: mathematics	9548	487.113	86.546	486.69	139.44	756.2	-0.07	-0.068	0.886
Plausible value 5: science	9548	508.775	98.938	510.525	99.25	831.26	-0.101	-0.211	1.013
Plausible value 5: reading	9548	490.434	95.309	493.035	148.04	765.81	-0.161	-0.155	0.975
Plausible value 5: mathematics	9548	487.001	86.152	487	170.44	801.3	-0.049	-0.177	0.882
Student level: ESCS	9352	0.18	0.774	0.163	-3.427	2.902	-0.018	0.017	0.008
School level: ESCS	9548	0.175	0.372	0.131	-0.771	1.295	0.313	-0.274	0.004
Student level: Migrant background second generation	9304	0.035	0.183	0	0	1	5.1	24.014	0.002
School level: % Migrant background second generation	9548	0.034	0.087	0	0	0.571	3.967	16.937	0.001
Language at home = Test language	9481	0.925	0.263	1	0	1	-3.23	8.432	0.003
Student level: School disciplinary climate (WLE)	9430	0.083	1.025	0.026	-2.809	1.838	-0.532	0.222	0.011
School level: Mean disciplinary climate	9548	0.081	0.38	0.088	-1.213	1.191	-0.236	0.335	0.004
School autonomy	8997	10.174	3.87	10	-6	20	-0.469	0.209	0.041
Educational leadership	8997	0.728	0.156	0.75	0.167	1	-0.817	0.525	0.002
Accountability	8847	7.021	1.232	7	2	8	-1.414	1.7	0.013

Table B3: Descriptive statistics of the independent and dependent variables used for the regression models for UK in 2015. Own calculations.

Summary UK 2015	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	11046	506.01	97.062	507.498	175.218	860.276	0.019	-0.416	0.924
Plausible value 1: reading	11046	496.858	93.242	498.02	163.49	846.678	-0.068	-0.143	0.887
Plausible value 1: mathematics	11046	489.293	88.654	491.127	109.656	790.172	-0.064	-0.18	0.844
Plausible value 2: science	11046	505.073	98.415	506.193	156.994	846.24	-0.016	-0.423	0.936
Plausible value 2: reading	11046	497.248	92.702	497.675	158.655	848.507	-0.055	-0.124	0.882
Plausible value 2: mathematics	11046	491.575	88.102	493.925	130.729	852.882	-0.083	-0.086	0.838
Plausible value 3: science	11046	505.829	97.665	506.948	105.751	840.99	-0.004	-0.396	0.929
Plausible value 3: reading	11046	493.926	93.239	495.33	73.188	814.148	-0.098	-0.119	0.887
Plausible value 3: mathematics	11046	491.172	86.314	492.967	47.62	790.529	-0.089	-0.078	0.821
Plausible value 4: science	11046	505.725	97.386	506.395	155.025	800.353	0.001	-0.464	0.927
Plausible value 4: reading	11046	494.067	92.587	494.187	132.515	798.514	-0.049	-0.194	0.881
Plausible value 4: mathematics	11046	492.963	85.938	493.944	126.812	786.49	-0.077	-0.125	0.818
Plausible value 5: science	11046	505.752	97.483	507.309	152.891	816.635	0.008	-0.41	0.928
Plausible value 5: reading	11046	498.062	94.013	498.856	178.794	828.337	-0.041	-0.206	0.895
Plausible value 5: mathematics	11046	490.987	86.878	492.219	96.374	810.008	-0.062	-0.042	0.827
Plausible value 6: science	11046	505.86	97.148	506.653	186.369	937.387	-0.004	-0.407	0.924
Plausible value 6: reading	11046	496.056	92.843	497.996	115.04	847.918	-0.078	-0.161	0.883
Plausible value 6: mathematics	11046	491.87	87.7	494.055	110.333	807.461	-0.092	-0.072	0.834
Plausible value 7: science	11046	505.33	97.196	507.106	160.433	813.822	0.007	-0.412	0.925
Plausible value 7: reading	11046	496.709	92.815	497.592	129.737	795.036	-0.041	-0.152	0.883
Plausible value 7: mathematics	11046	494.838	86.28	496.802	112.653	839.195	-0.103	-0.065	0.821
Plausible value 8: science	11046	504.327	97.151	504.795	169.13	843.229	0.013	-0.396	0.924
Plausible value 8: reading	11046	494.393	92.922	495.344	93.922	798.983	-0.092	-0.155	0.884
Plausible value 8: mathematics	11046	493.138	87.095	496.11	113.363	808.305	-0.136	-0.056	0.829
Plausible value 9: science	11046	504.541	97.244	506.123	141.394	865.208	0.022	-0.356	0.925
Plausible value 9: reading	11046	499.955	93.501	499.659	151.949	851.716	-0.026	-0.131	0.89
Plausible value 9: mathematics	11046	494.371	87.931	496.092	63.876	786.388	-0.078	-0.102	0.837
Plausible value 10: science	11046	505.565	97.119	506.708	166.046	842.296	0.003	-0.401	0.924
Plausible value 10: reading	11046	495.724	92.67	496.15	166.755	837.953	-0.034	-0.123	0.882
Plausible value 10: mathematics	11046	492.059	86.696	492.568	175.843	820.052	-0.025	-0.138	0.825

Table B3 (continued).

Summary UK 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Student level: ESCS	10536	0.22	0.859	0.264	-6.656	3.959	-0.213	-0.12	0.008
School level: ESCS	10997	0.211	0.4	0.19	-0.684	1.263	0.31	-0.55	0.004
Student level: Migrant background second generation	10435	0.056	0.23	0	0	1	3.867	12.955	0.002
School level: % Migrant background second generation	10973	0.056	0.112	0	0	0.679	2.817	8.104	0.001
Language at home = Test language	10786	0.913	0.282	1	0	1	-2.929	6.58	0.003
Student level: School disciplinary climate (WLE)	10075	-0.087	1.035	0.004	-2.416	1.884	-0.18	-0.17	0.01
School level: Mean disciplinary climate	10973	-0.086	0.406	-0.091	-1.069	1.022	0.167	-0.308	0.004
School autonomy	8809	0.909	0.192	1	0	1	-3.688	14.413	0.002
Educational leadership	8486	0.71	1.188	0.591	-1.658	4.43	1.17	1.799	0.013
Accountability	8113	4.903	1.327	5	0	6	-1.607	2.854	0.015

Table B4: Descriptive statistics of the independent and dependent variables used for the regression models for Germany in 2000. Own calculations.

Summary Germany 2000	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	2855	495.549	101.681	501.1	146.24	794.31	-0.201	-0.279	1.903
Plausible value 1: reading	5073	497.679	103.687	504.84	107.56	779.38	-0.364	-0.019	1.456
Plausible value 1: mathematics	2830	500.095	99.701	505.28	77.15	787.09	-0.265	-0.078	1.874
Plausible value 2: science	2855	496.968	100.149	500.23	168.86	823.17	-0.13	-0.256	1.874
Plausible value 2: reading	5073	498.368	103.117	505.68	95.74	789.26	-0.356	-0.05	1.448
Plausible value 2: mathematics	2830	500.11	99.15	507.61	117.84	770.3	-0.288	-0.122	1.864
Plausible value 3: science	2855	496.752	100.132	499.49	142.98	807.84	-0.213	-0.261	1.874
Plausible value 3: reading	5073	497.903	103.714	505.31	81.2	825.9	-0.352	0.017	1.456
Plausible value 3: mathematics	2830	499.42	99.68	503.79	149.89	780.23	-0.227	-0.17	1.874
Plausible value 4: science	2855	496.465	100.76	501.14	155.64	784.35	-0.185	-0.337	1.886
Plausible value 4: reading	5073	498.267	103.96	505.75	76.65	780.29	-0.382	0.065	1.46
Plausible value 4: mathematics	2830	500.311	100.098	506.395	136.15	843.57	-0.266	-0.097	1.882
Plausible value 5: science	2855	496.303	100.44	499.85	169.14	795.21	-0.174	-0.351	1.88
Plausible value 5: reading	5073	497.693	104.5	504.88	91.32	791.08	-0.367	0.045	1.467
Plausible value 5: mathematics	2830	500.238	100.551	507.32	32.88	789.38	-0.281	-0.067	1.89
Student level: ESCS	5019	0.268	0.91	0.192	-4.911	3.257	0.005	0.058	0.013
School level: ESCS	5073	0.265	0.48	0.239	-1.502	1.389	0.005	-0.576	0.007
Student level: Migrant background second generation	5073	0.109	0.312	0	0	1	2.505	4.277	0.004
School level: % Migrant background second generation	5073	0.109	0.106	0.077	0	0.667	1.419	2.588	0.001
Language at home = Test language	4617	0.929	0.256	1	0	1	-3.351	9.233	0.004
Student level: School disciplinary climate (WLE)	5073	1.422	12.218	-0.19	-2.92	97	7.639	56.791	0.172
School level: Mean disciplinary climate	5073	1.422	9.642	-0.075	-1.168	97	9.163	86.836	0.135
School autonomy	4630	-0.977	0.534	-0.84	-3.22	1.72	0.03	4.664	0.008
Educational leadership	4630	0.009	0.033	0	0	0.2	3.826	15.01	0
Accountability	4604	3.085	0.992	3	0	6	0.324	0.096	0.015

Table B5: Descriptive statistics of the independent and dependent variables used for the regression models for Germany in 2009. Own calculations.

Summary Germany 2009	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	4979	518.929	101.38	524.75	146.16	810.74	-0.241	-0.279	1.437
Plausible value 1: reading	4979	495.812	95.648	504.06	128.96	764.77	-0.308	-0.247	1.356
Plausible value 1: mathematics	4979	511.035	98.531	513.56	159.53	825.14	-0.16	-0.344	1.396
Plausible value 2: science	4979	518.776	100.466	523.82	134.97	827.81	-0.23	-0.3	1.424
Plausible value 2: reading	4979	495.39	95.538	503.66	149.8	756.57	-0.316	-0.254	1.354
Plausible value 2: mathematics	4979	511.077	98.212	516.29	186.02	816.88	-0.138	-0.359	1.392
Plausible value 3: science	4979	519.479	101.213	525.96	173.2	809.81	-0.248	-0.32	1.434
Plausible value 3: reading	4979	495.552	96.357	503.1	108.92	749.9	-0.296	-0.291	1.366
Plausible value 3: mathematics	4979	511.982	98.359	515.51	127.68	809.17	-0.148	-0.378	1.394
Plausible value 4: science	4979	519.269	101.023	526.61	180.75	826.59	-0.243	-0.273	1.432
Plausible value 4: reading	4979	495.663	95.577	504.98	141.79	763.4	-0.294	-0.27	1.355
Plausible value 4: mathematics	4979	511.641	98.532	515.51	163.43	836.04	-0.176	-0.316	1.396
Plausible value 5: science	4979	519.542	102.203	525.68	136.93	828.55	-0.204	-0.344	1.448
Plausible value 5: reading	4979	496.376	96.496	503.18	176.58	781.91	-0.253	-0.376	1.368
Plausible value 5: mathematics	4979	512.113	98.978	515.9	179.71	827.47	-0.136	-0.383	1.403
Student level: ESCS	4561	0.173	0.907	0.132	-4.81	3.157	-0.102	0.313	0.013
School level: ESCS	4979	0.17	0.48	0.164	-1.035	1.097	-0.056	-0.686	0.007
Student level: Migrant background second generation	4518	0.116	0.321	0	0	1	2.391	3.718	0.005
School level: % Migrant background second generation	4979	0.116	0.13	0.083	0	0.81	1.742	4.04	0.002
Language at home = Test language	4645	0.83	0.376	1	0	1	-1.758	1.09	0.006
Student level: School disciplinary climate (WLE)	4443	0.252	0.987	0.351	-2.809	1.838	-0.666	0.657	0.015
School level: Mean disciplinary climate	4800	0.254	0.361	0.266	-1.261	1.58	-0.347	1.342	0.005
School autonomy	4701	3.158	4.06	3	-9	20	0.34	1.208	0.059
Educational leadership	4701	0.445	0.176	0.417	0	0.917	-0.062	-0.151	0.003
Accountability	4500	4.265	1.594	4	0	8	0.172	-0.69	0.024

Table B6: Descriptive statistics of the independent and dependent variables used for the regression models for Germany in 2015. Own calculations.

Summary Germany 2015	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	6504	512.269	98.545	515.968	175.599	814.68	-0.098	-0.321	1.222
Plausible value 1: reading	6504	512.35	98.682	518.135	171.756	808.879	-0.259	-0.093	1.224
Plausible value 1: mathematics	6504	507.994	88.877	509.608	161.574	803.717	-0.139	-0.029	1.102
Plausible value 2: science	6504	511.648	99.446	513.99	131.415	833.682	-0.108	-0.302	1.233
Plausible value 2: reading	6504	513.411	99.149	521.17	135.72	796.595	-0.279	-0.124	1.229
Plausible value 2: mathematics	6504	510.55	88.988	513.489	206.821	794.11	-0.118	-0.138	1.103
Plausible value 3: science	6504	511.812	98.423	515.585	183.505	810.632	-0.091	-0.345	1.22
Plausible value 3: reading	6504	510.569	99.375	517.308	138.637	797.475	-0.28	-0.072	1.232
Plausible value 3: mathematics	6504	509.17	87.815	511.604	195.138	789.649	-0.118	-0.195	1.089
Plausible value 4: science	6504	511.637	99.362	514.994	156.365	826.693	-0.099	-0.314	1.232
Plausible value 4: reading	6504	511.48	99.867	518.09	167.892	854.437	-0.277	-0.171	1.238
Plausible value 4: mathematics	6504	509.473	89.678	511.857	176.137	807.894	-0.071	-0.174	1.112
Plausible value 5: science	6504	512.277	99.472	514.64	161.822	817.277	-0.088	-0.349	1.233
Plausible value 5: reading	6504	512.345	99.526	518.798	166.293	809.115	-0.293	-0.153	1.234
Plausible value 5: mathematics	6504	509.082	88.81	510.951	223.043	800.311	-0.057	-0.218	1.101
Plausible value 6: science	6504	512.989	98.981	516.198	115.876	816.741	-0.101	-0.341	1.227
Plausible value 6: reading	6504	512.941	101.122	519.157	60.263	822.427	-0.294	-0.07	1.254
Plausible value 6: mathematics	6504	505.792	89.874	507.042	159.482	799.123	-0.106	-0.075	1.114
Plausible value 7: science	6504	513.222	99.325	515.62	145.715	834.441	-0.069	-0.303	1.232
Plausible value 7: reading	6504	511.67	99.522	518.523	123.195	809.445	-0.264	-0.147	1.234
Plausible value 7: mathematics	6504	510.139	89.005	510.726	194.582	791.261	-0.076	-0.183	1.104
Plausible value 8: science	6504	511.905	100.064	515.468	180.128	845.713	-0.074	-0.284	1.241
Plausible value 8: reading	6504	512.07	99.521	520.565	163.134	847.261	-0.28	-0.089	1.234
Plausible value 8: mathematics	6504	507.687	89.013	510.207	197.464	845.993	-0.08	-0.096	1.104
Plausible value 9: science	6504	512.208	100.109	515.342	157.886	852.759	-0.093	-0.334	1.241
Plausible value 9: reading	6504	512.486	99.214	519.01	112.564	802.26	-0.288	-0.14	1.23
Plausible value 9: mathematics	6504	507.911	88.747	509.99	187.785	810.71	-0.122	-0.154	1.1
Plausible value 10: science	6504	512.104	98.609	514.118	212.578	836.05	-0.06	-0.312	1.223
Plausible value 10: reading	6504	512.243	100.57	519.06	140.087	837.863	-0.238	-0.141	1.247
Plausible value 10: mathematics	6504	509.628	89.305	511.233	191.304	808.795	-0.119	-0.191	1.107

Table B6 (continued).

Summary Germany 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Student level: ESCS	5630	0.139	0.945	0.094	-4.161	3.688	-0.113	-0.33	0.013
School level: ESCS	6497	0.12	0.491	0.087	-1.961	1.134	-0.037	-0.416	0.006
Student level: Migrant background second generation	5691	0.132	0.339	0	0	1	2.172	2.718	0.004
School level: % Migrant background second generation	6497	0.137	0.135	0.1	0	0.778	1.37	2.18	0.002
Language at home = Test language	5794	0.881	0.324	1	0	1	-2.349	3.518	0.004
Student level: School disciplinary climate (WLE)	4871	0.055	0.939	0.004	-2.416	1.884	-0.167	-0.002	0.013
School level: Mean disciplinary climate	6437	0.022	0.354	0.036	-1.614	1.206	-0.692	1.201	0.004
School autonomy	5010	0.625	0.122	0.67	0	1	-0.672	3.136	0.002
Educational leadership	4954	-0.146	0.912	-0.222	-2.204	4.43	1.391	5.112	0.013
Accountability	3827	1.297	1.768	0	0	6	1.121	-0.07	0.029

Table B7: Descriptive statistics of the independent and dependent variables used for the regression models for Sweden in 2000. Own calculations.

Summary Sweden 2000	n	mean	sd	median	min	max	skew	kurtosis	se
Plausible value 1: science	2470	511.343	93.96	517.215	147.14	804.75	-0.164	-0.134	1.891
Plausible value 1: reading	4468	515.584	92.052	522.35	165.13	787.57	-0.309	-0.07	1.377
Plausible value 1: mathematics	2496	509.733	93.31	513.02	167.35	800.69	-0.251	0.179	1.868
Plausible value 2: science	2470	511.709	93.559	518.26	164.28	778.97	-0.194	-0.245	1.883
Plausible value 2: reading	4468	516.906	92.901	523.83	160.59	770.72	-0.326	0.009	1.39
Plausible value 2: mathematics	2496	509.248	92.842	513.86	164.8	795.23	-0.284	0.061	1.858
Plausible value 3: science	2470	512.156	92.602	517.295	198.9	767.77	-0.14	-0.178	1.863
Plausible value 3: reading	4468	516.871	92.39	523.395	157.63	818.35	-0.297	-0.034	1.382
Plausible value 3: mathematics	2496	510.682	93.417	515.55	138.21	793.71	-0.258	0.05	1.87
Plausible value 4: science	2470	512.435	93.362	514.67	127.3	791.56	-0.101	-0.259	1.879
Plausible value 4: reading	4468	516.74	91.957	523.395	172.21	805.61	-0.274	-0.098	1.376
Plausible value 4: mathematics	2496	509.269	92.588	513.09	99.28	797.55	-0.226	0.075	1.853
Plausible value 5: science	2470	512.759	93.484	517.71	160.67	881.43	-0.115	-0.119	1.881
Plausible value 5: reading	4468	516.309	91.762	522.15	111.51	860.24	-0.328	0.093	1.373
Plausible value 5: mathematics	2496	509.379	93.901	514.48	151.18	822.74	-0.294	0.245	1.88
Student level: ESCS	4436	0.348	0.804	0.365	-3.719	2.63	-0.289	0.248	0.012
School level: ESCS	4468	0.348	0.308	0.297	-0.525	1.4	0.817	0.784	0.005
Student level: Migrant background second generation	4468	0.145	0.352	0	16	99	0.785	0.077	0.273
School level: % Migrant background second generation	4468	0.145	0.115	0.125	30	67.417	0.861	0.822	0.093
Language at home = Test language	4289	0.932	0.252	1	0	1	2.021	2.083	0.005
Student level: School disciplinary climate (WLE)	4468	1.033	9.046	0.1	0	0.714	1.799	4.635	0.002
School level: Mean disciplinary climate	4468	1.033	1.924	0.279	0	1	-3.421	9.706	0.004
School autonomy	4468	0.714	0.821	0.98	-2.92	97	10.419	107.574	0.135
Educational leadership	4468	0.004	0.025	0	-0.9	14.889	2.672	8.776	0.029
Accountability	4378	3.805	1.114	4	-1.71	1.72	-0.396	-0.563	0.012

Table B8: Descriptive statistics of the independent and dependent variables used for the regression models for Sweden in 2009. Own calculations.

Summary Sweden 2009	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	4567	495.601	99.774	498.64	86.48	782.86	-0.186	0.005	1.476
Plausible value 1: reading	4567	498.543	98.394	502.68	127.44	807.17	-0.309	0.156	1.456
Plausible value 1: mathematics	4567	494.865	93.763	497.59	122.53	786.58	-0.142	0.024	1.387
Plausible value 2: science	4567	496.087	99.413	496.87	70.44	793.58	-0.224	0.154	1.471
Plausible value 2: reading	4567	498.231	98.79	502.52	109.32	789.38	-0.333	0.158	1.462
Plausible value 2: mathematics	4567	496.077	93.768	497.9	126.74	792.03	-0.162	0.023	1.388
Plausible value 3: science	4567	496.017	98.474	499.57	111.75	820.35	-0.225	0.074	1.457
Plausible value 3: reading	4567	497.598	97.936	502.68	50.32	786.91	-0.338	0.281	1.449
Plausible value 3: mathematics	4567	494.989	92.979	496.74	153.3	803.79	-0.163	0.038	1.376
Plausible value 4: science	4567	495.483	99.837	497.71	76.97	819.41	-0.232	0.104	1.477
Plausible value 4: reading	4567	498.357	98.466	503.26	68.43	795.1	-0.327	0.275	1.457
Plausible value 4: mathematics	4567	495.192	94.229	496.74	112.8	796.78	-0.176	0.081	1.394
Plausible value 5: science	4567	495.685	100.051	497.71	106.06	800.48	-0.197	0.083	1.48
Plausible value 5: reading	4567	498.824	98.803	502.38	50.23	784.05	-0.308	0.236	1.462
Plausible value 5: mathematics	4567	494.812	93.716	496.74	111.94	782.68	-0.157	0.005	1.387
Student level: ESCS	4512	0.336	0.812	0.354	-6.037	2.987	-0.391	1.107	0.012
School level: ESCS	4567	0.334	0.345	0.307	-1.231	1.645	0.428	0.757	0.005
Student level: Migrant background second generation	4500	0.076	0.266	0	16	88	0.492	-0.712	0.252
School level: % Migrant background second generation	4567	0.077	0.107	0.038	23	78.5	0.52	0.144	0.105
Language at home = Test language	4548	0.88	0.325	1	0	1	3.187	8.159	0.004
Student level: School disciplinary climate (WLE)	4511	-0.035	0.903	0.026	0	1	2.817	12.267	0.002
School level: Mean disciplinary climate	4567	-0.035	0.344	-0.046	0	1	-2.34	3.478	0.005
School autonomy	4567	9.39	3.748	10	-2.809	1.838	-0.269	0.412	0.013
Educational leadership	4567	0.711	0.15	0.75	-1.075	1.838	0.197	0.896	0.005
Accountability	4435	5.27	1.498	5	0	20	-0.113	-0.431	0.055

Table B9: Descriptive statistics of the independent and dependent variables used for the regression models for Sweden in 2015. Own calculations.

Summary Sweden 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Plausible value 1: science	5458	492.541	102.113	494.812	142.748	845.611	-0.087	-0.271	1.382
Plausible value 1: reading	5458	499.757	101.913	505.914	95.962	826.607	-0.355	0.038	1.379
Plausible value 1: mathematics	5458	493.926	87.833	495.401	197.292	771.176	-0.11	-0.124	1.189
Plausible value 2: science	5458	493.221	102.984	496.417	160.248	836.631	-0.093	-0.286	1.394
Plausible value 2: reading	5458	499.98	101.821	507.858	111.79	852.872	-0.31	-0.043	1.378
Plausible value 2: mathematics	5458	492.432	89.969	495.468	136.394	766.22	-0.165	-0.098	1.218
Plausible value 3: science	5458	492.713	102.114	494.617	163.481	819.942	-0.067	-0.331	1.382
Plausible value 3: reading	5458	500.4	100.896	506.122	156.032	787.924	-0.26	-0.15	1.366
Plausible value 3: mathematics	5458	494.844	88.978	498.896	163.703	778.852	-0.175	-0.206	1.204
Plausible value 4: science	5458	492.593	102.053	494.864	106.739	826.56	-0.084	-0.303	1.381
Plausible value 4: reading	5458	500.809	100.413	507.716	126.705	794.526	-0.287	-0.077	1.359
Plausible value 4: mathematics	5458	493.337	88.854	493.886	87.785	823.08	-0.124	-0.067	1.203
Plausible value 5: science	5458	493.796	101.269	496.229	157.275	833.764	-0.061	-0.325	1.371
Plausible value 5: reading	5458	500.035	101.03	506.566	106.151	782.244	-0.304	-0.066	1.368
Plausible value 5: mathematics	5458	493.639	88.736	496.471	164.697	801.512	-0.096	-0.154	1.201
Plausible value 6: science	5458	493.755	101.865	497.245	109.772	869.5	-0.077	-0.289	1.379
Plausible value 6: reading	5458	500.962	100.79	508.274	64.858	799.448	-0.353	-0.004	1.364
Plausible value 6: mathematics	5458	493.669	91.627	495.804	153.494	792.57	-0.154	-0.088	1.24
Plausible value 7: science	5458	493.405	101.858	495.662	142.177	830.544	-0.075	-0.357	1.379
Plausible value 7: reading	5458	499.774	102.049	506.101	128.061	774.474	-0.306	-0.166	1.381
Plausible value 7: mathematics	5458	492.759	89.502	496.082	162.081	753.642	-0.159	-0.137	1.211
Plausible value 8: science	5458	492.175	101.472	494.826	159.952	810.203	-0.08	-0.343	1.373
Plausible value 8: reading	5458	501.516	102.32	508.729	102.366	800.445	-0.315	-0.061	1.385
Plausible value 8: mathematics	5458	493.003	89.115	496.348	157.635	772.031	-0.157	-0.11	1.206
Plausible value 9: science	5458	492.061	102.925	493.612	141.469	824.222	-0.064	-0.332	1.393
Plausible value 9: reading	5458	499.475	101.84	505.569	122.261	804.143	-0.279	-0.053	1.378
Plausible value 9: mathematics	5458	494.565	90.593	496.073	192.195	789.379	-0.12	-0.212	1.226
Plausible value 10: science	5458	493.633	101.469	497.012	173.646	821.044	-0.082	-0.343	1.373
Plausible value 10: reading	5458	499.172	101.656	505.434	116.66	834.854	-0.323	-0.01	1.376
Plausible value 10: mathematics	5458	493.828	88.459	496.275	160.864	791.479	-0.124	-0.179	1.197

Table B9 (continued).

Summary Sweden 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Student level: ESCS	5313	0.338	0.819	0.458	-3.955	3.062	-0.597	0.448	0.011
School level: ESCS	5458	0.335	0.335	0.324	-1.267	1.29	0.116	-0.351	0.005
Student level: Migrant background second generation	5283	0.095	0.294	0	12	89	0.183	-1.386	0.322
School level: % Migrant background second generation	5458	0.096	0.121	0.057	22	82	0.318	-0.35	0.125
Language at home = Test language	5375	0.852	0.356	1	0	1	2.758	5.605	0.004
Student level: School disciplinary climate (WLE)	5065	0.015	0.951	0.004	0	0.727	2.043	5.102	0.002
School level: Mean disciplinary climate	5456	0.007	0.384	0.007	0	1	-1.977	1.908	0.005
School autonomy	5458	0.878	0.123	0.92	-2.416	1.884	-0.1	0.044	0.013
Educational leadership	5443	0.108	0.829	0.004	-1.55	1.884	0.165	0.28	0.005
Accountability	5282	3.601	1.102	4	0.33	1	-1.05	1.307	0.002

Table B10: Descriptive statistics of the independent and dependent variables used for the regression models for Finland in 2000. Own calculations.

Summary Finland 2000	**n**	**mean**	**sd**	**median**	**min**	**max**	**skew**	**kurtosis**	**se**
Plausible value 1: science	2739	538.783	86.267	540.98	226.16	793.93	-0.15	-0.093	1.648
Plausible value 1: reading	4908	548.775	87.414	554.185	160.28	838.42	-0.331	0.21	1.248
Plausible value 1: mathematics	2726	538.196	79.212	540.245	114.26	775.39	-0.27	0.384	1.517
Plausible value 2: science	2739	538.291	85.552	540.51	166.99	786.15	-0.175	-0.001	1.635
Plausible value 2: reading	4908	548.101	87.871	553.06	153	850.28	-0.35	0.214	1.254
Plausible value 2: mathematics	2726	536.373	79.962	539.66	263.29	778.47	-0.215	0.006	1.532
Plausible value 3: science	2739	538.875	86.108	541.73	234.64	834.86	-0.176	-0.064	1.645
Plausible value 3: reading	4908	548.048	86.614	553.655	165.73	840.44	-0.338	0.255	1.236
Plausible value 3: mathematics	2726	536.837	79.623	538.21	163.53	786.11	-0.235	0.325	1.525
Plausible value 4: science	2739	537.468	85.049	539.61	179.62	827.3	-0.207	0.071	1.625
Plausible value 4: reading	4908	547.562	87.165	553.06	119.37	817.52	-0.379	0.414	1.244
Plausible value 4: mathematics	2726	537.728	79.475	540.43	178.37	789.72	-0.234	0.216	1.522
Plausible value 5: science	2739	539.305	85.696	542.32	229.61	800.24	-0.177	0.074	1.637
Plausible value 5: reading	4908	548.242	87.586	554.185	168.46	855.93	-0.315	0.142	1.25
Plausible value 5: mathematics	2726	536.806	79.959	538.62	230.47	775.98	-0.19	0.15	1.531
Student level: ESCS	4857	0.042	0.939	-0.001	-3.261	3.094	0.058	-0.689	0.013
School level: ESCS	4908	0.043	0.37	0.019	-1.018	1.233	0.506	0.365	0.005
Student level: Migrant background second generation	4908	0.022	0.148	0	16	99	0.709	-0.154	0.263
School level: % Migrant background second generation	4908	0.022	0.029	0	27.8	65	0.382	-0.217	0.1
Language at home = Test language	4674	0.986	0.117	1	0	1	6.451	39.624	0.002
Student level: School disciplinary climate (WLE)	4908	1.105	9.574	0.1	0	0.152	1.51	2.52	0
School level: Mean disciplinary climate	4908	1.105	2.037	0.292	0	1	-8.299	66.892	0.002
School autonomy	4908	-0.439	0.511	-0.57	-2.92	97	9.822	95.419	0.137
Educational leadership	4729	0.006	0.029	0	-0.911	11.526	2.582	7.771	0.029
Accountability	4823	3.766	1.211	4	-1.71	1.72	0.073	0.22	0.007

Table B11: Descriptive statistics of the independent and dependent variables used for the regression models for Finland in 2009. Own calculations.

Summary Finland 2009	n	mean	sd	median	min	max	skew	kurtosis	se
Plausible value 1: science	5810	548.6	89.299	551.79	229.9	856.53	-0.199	-0.07	1.172
Plausible value 1: reading	5810	530.977	86.756	536.405	146.44	802.88	-0.337	-0.018	1.138
Plausible value 1: mathematics	5810	537.481	81.809	540.05	217.1	785.33	-0.157	-0.06	1.073
Plausible value 2: science	5810	549.385	89.193	553.66	186.26	836.2	-0.228	0.044	1.17
Plausible value 2: reading	5810	531.492	87.278	536.84	180.91	786.68	-0.334	-0.01	1.145
Plausible value 2: mathematics	5810	538.042	82.089	540.82	200.82	811.04	-0.201	0.036	1.077
Plausible value 3: science	5810	549.595	89.634	552.72	190.64	839	-0.19	-0.015	1.176
Plausible value 3: reading	5810	531.587	86.778	538.43	134.55	810.83	-0.374	0.05	1.138
Plausible value 3: mathematics	5810	537.784	82.623	540.05	116.3	845.31	-0.198	0.153	1.084
Plausible value 4: science	5810	549.055	89.267	553.38	237.54	811.3	-0.252	0.003	1.171
Plausible value 4: reading	5810	531.537	86.795	538.28	146.44	775.87	-0.35	0.005	1.139
Plausible value 4: mathematics	5810	538.044	82.288	540.55	227.22	784.55	-0.212	-0.056	1.08
Plausible value 5: science	5810	549.001	89.502	553.38	174.79	842.73	-0.209	-0.003	1.174
Plausible value 5: reading	5810	531.477	86.674	537.29	172.25	827.98	-0.349	0.061	1.137
Plausible value 5: mathematics	5810	537.931	82.36	540.05	202.38	829.73	-0.2	0.021	1.081
Student level: ESCS	5776	0.413	0.78	0.441	-3.562	3.056	-0.319	0.079	0.01
School level: ESCS	5810	0.413	0.312	0.35	-0.324	1.361	0.538	-0.194	0.004
Student level: Migrant background second generation	5754	0.01	0.101	0	16	90	0.357	-1.065	0.237
School level: % Migrant background second generation	5810	0.01	0.021	0	26.5	79	0.475	-0.312	0.091
Language at home = Test language	5802	0.783	0.412	1	0	1	9.72	92.503	0.001
Student level: School disciplinary climate (WLE)	5774	-0.267	0.944	-0.287	0	0.133	2.354	6.63	0
School level: Mean disciplinary climate	5810	-0.267	0.369	-0.291	0	1	-1.372	-0.119	0.005
School autonomy	5810	4.076	4.28	4	-2.809	1.838	-0.256	0.189	0.012
Educational leadership	5810	0.595	0.201	0.583	-2.363	1.838	0.058	2.139	0.005
Accountability	5774	4.348	1.66	4	-6	17	0.116	-0.256	0.056

Table B12: Descriptive statistics of the independent and dependent variables used for the regression models for Finland in 2015. Own calculations.

Summary Finland 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Plausible value 1: science	5882	531.985	96.344	536.668	199.678	852.902	-0.43	0.298	1.193
Plausible value 1: reading	5882	527.861	91.47	535.014	73.377	778.198	-0.145	-0.082	1.069
Plausible value 1: mathematics	5882	511.283	81.981	513.55	208.486	755.785	-0.199	-0.198	1.242
Plausible value 2: science	5882	532.078	95.262	536.094	168.541	824.971	-0.453	0.317	1.211
Plausible value 2: reading	5882	527.673	92.897	534.915	46.927	810.947	-0.189	-0.012	1.051
Plausible value 2: mathematics	5882	512.746	80.631	514.805	214.305	791.821	-0.184	-0.118	1.244
Plausible value 3: science	5882	531.484	95.431	535.919	205.719	857.102	-0.421	0.262	1.228
Plausible value 3: reading	5882	527.44	94.218	534.514	57.679	816.232	-0.189	0.076	1.073
Plausible value 3: mathematics	5882	509.746	82.314	512.558	132.36	791.673	-0.13	-0.208	1.244
Plausible value 4: science	5882	531.963	95.369	535.054	224.862	865.721	-0.392	0.204	1.225
Plausible value 4: reading	5882	527.499	93.918	534.593	116.678	816.582	-0.138	-0.008	1.069
Plausible value 4: mathematics	5882	512.518	82.023	514.772	194.299	787.899	-0.17	-0.201	1.238
Plausible value 5: science	5882	531.311	94.918	534.547	170.749	853.444	-0.437	0.377	1.217
Plausible value 5: reading	5882	527.339	93.331	535.068	96.893	852.137	-0.152	0.033	1.07
Plausible value 5: mathematics	5882	510.872	82.085	512.811	179.427	800.947	-0.173	-0.096	1.25
Plausible value 6: science	5882	531.748	95.889	537.25	150.889	877.32	-0.438	0.382	1.209
Plausible value 6: reading	5882	527.777	92.698	535.722	0	870.747	-0.153	-0.041	1.056
Plausible value 6: mathematics	5882	511.628	81.01	514.192	238.882	782.786	-0.211	-0.198	1.255
Plausible value 7: science	5882	530.824	96.275	536.888	194.602	824.409	-0.482	0.452	1.223
Plausible value 7: reading	5882	527.976	93.814	536.5	28.659	788.424	-0.147	0.046	1.06
Plausible value 7: mathematics	5882	512.232	81.261	514.454	192.71	775.529	-0.181	-0.147	1.246
Plausible value 8: science	5882	531.511	95.536	535.822	194.6	873.018	-0.436	0.359	1.209
Plausible value 8: reading	5882	527.099	92.689	534.882	46.421	778.992	-0.207	0.186	1.069
Plausible value 8: mathematics	5882	512.593	82.004	515.976	186.6	811.91	-0.178	-0.226	1.256
Plausible value 9: science	5882	531.354	96.307	535.806	217.444	827.727	-0.484	0.378	1.215
Plausible value 9: reading	5882	528.58	93.166	537.604	22.847	794.958	-0.19	0.025	1.06
Plausible value 9: mathematics	5882	513.188	81.277	515.963	170.265	777.944	-0.227	-0.189	1.247
Plausible value 10: science	5882	531.46	95.616	536.181	197.709	856.427	-0.399	0.161	1.231
Plausible value 10: reading	5882	527.224	94.386	534.73	125.688	807.257	-0.43	0.298	1.193
Plausible value 10: mathematics	5882	511.861	82.021	515.476	199.678	852.902	-0.145	-0.082	1.069

Table B12 (continued).

Summary Finland 2015	n	mean	sd	median	min	max	skew	kurtosis	se
Student level: ESCS	5812	0.259	0.748	0.312	-4.112	3.567	-0.299	-0.014	0.01
School level: ESCS	5882	0.258	0.293	0.237	-0.97	1.37	0.265	0.302	0.004
Student level: Migrant background second generation	5794	0.018	0.133	0	12	89	0.473	-1.27	0.31
School level: % Migrant background second generation	5882	0.018	0.04	0	24	75	0.434	-0.329	0.113
Language at home = Test language	5816	0.911	0.285	1	0	1	7.26	50.711	0.002
Student level: School disciplinary climate (WLE)	5672	-0.106	0.881	0.004	0	0.294	4.644	26.488	0.001
School level: Mean disciplinary climate	5882	-0.108	0.294	-0.123	0	1	-2.877	6.28	0.004
School autonomy	5842	0.748	0.161	0.75	-2.416	1.884	-0.062	0.412	0.012
Educational leadership	5770	-0.193	0.758	-0.181	-1.114	1.884	0.069	0.733	0.004
Accountability	5132	2.653	1.526	3	0	1	-1.376	4.464	0.002

Figure B1: Heatmap of the correlations among independent and dependent variables used in the regression models for UK in 2000. Own calculations.

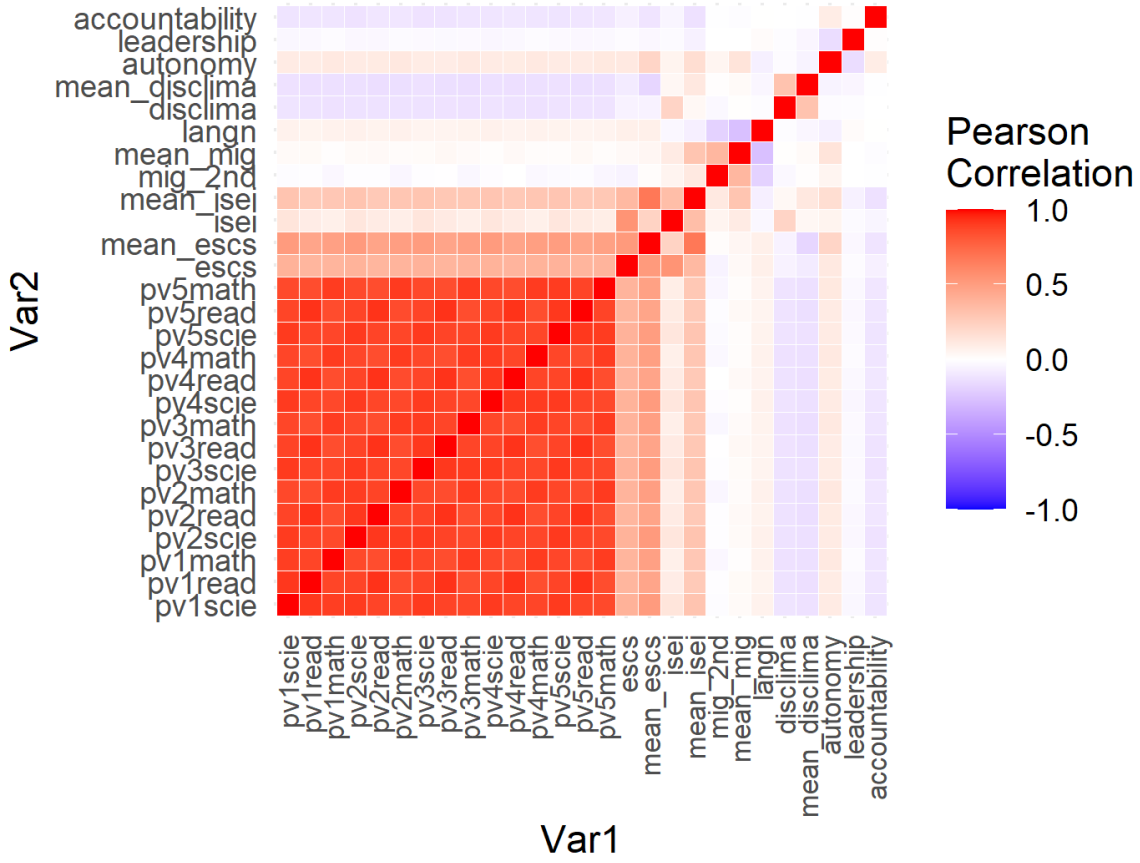


Figure B2: Heatmap of the correlations among independent and dependent variables used in the regression models for UK in 2009. Own calculations.

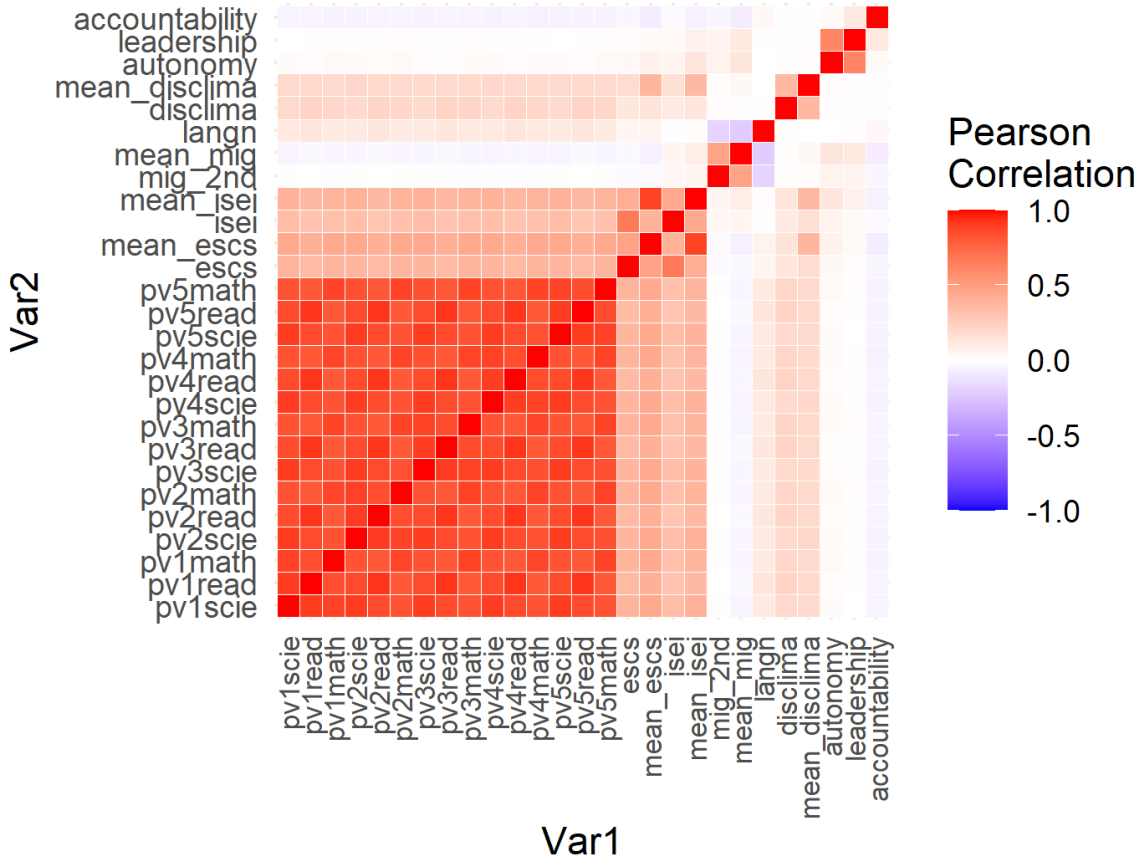


Figure B3: Heatmap of the correlations among independent and dependent variables used in the regression models for UK in 2015. Own calculations.

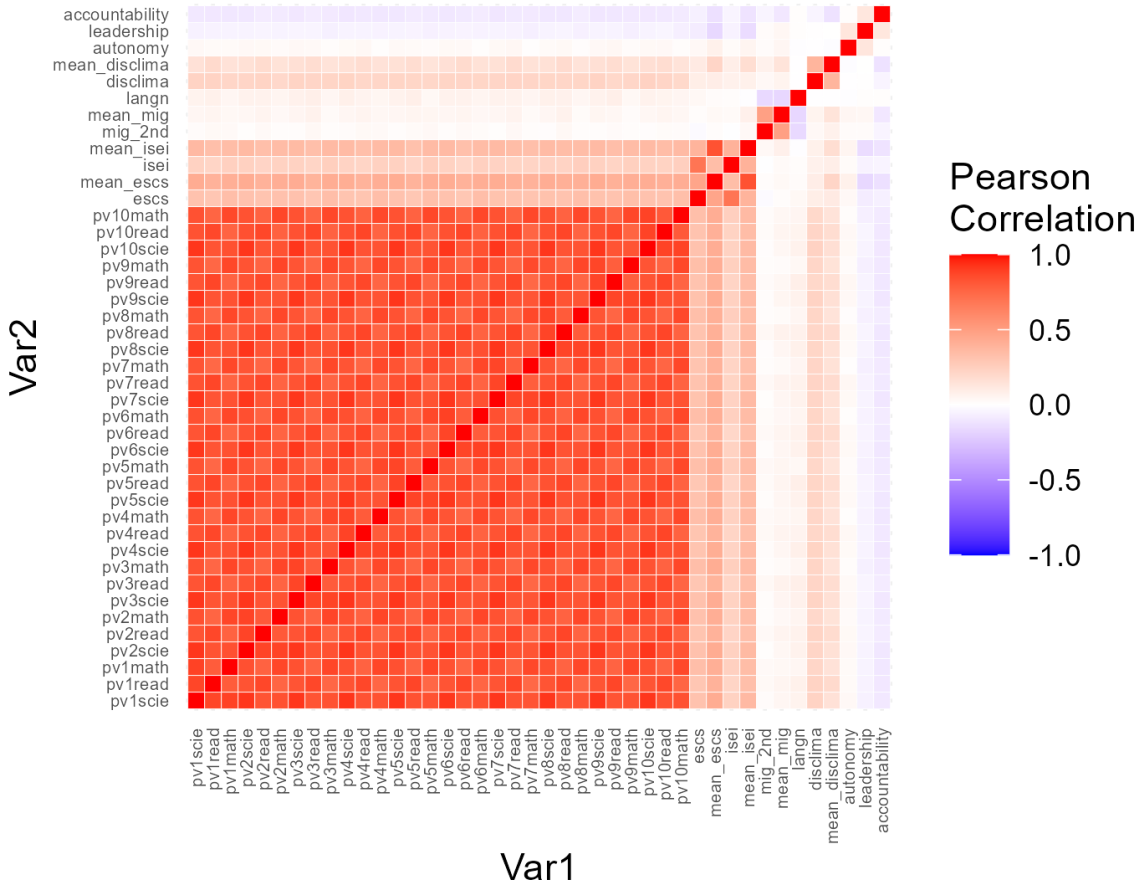


Figure B4: Heatmap of the correlations among independent and dependent variables used in the regression models for Germany in 2000. Own calculations.

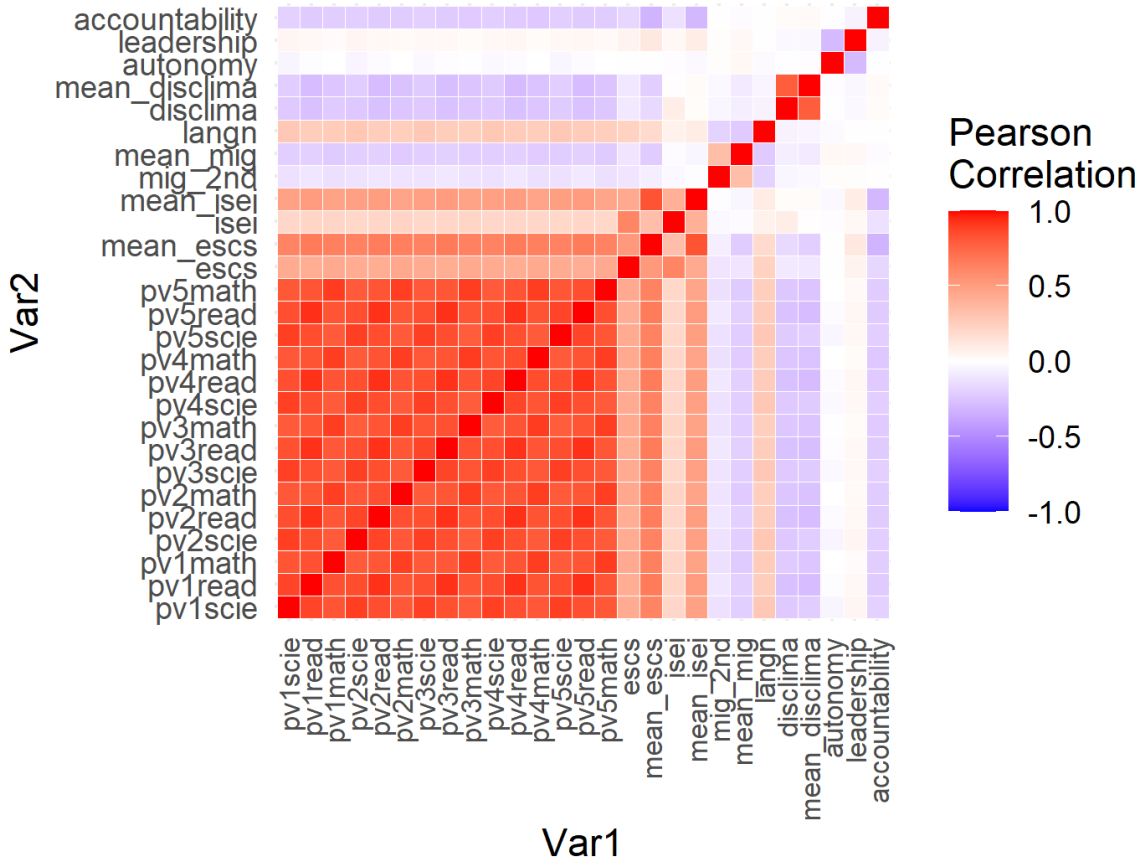


Figure B5: Heatmap of the correlations among independent and dependent variables used in the regression models for Germany in 2009. Own calculations.

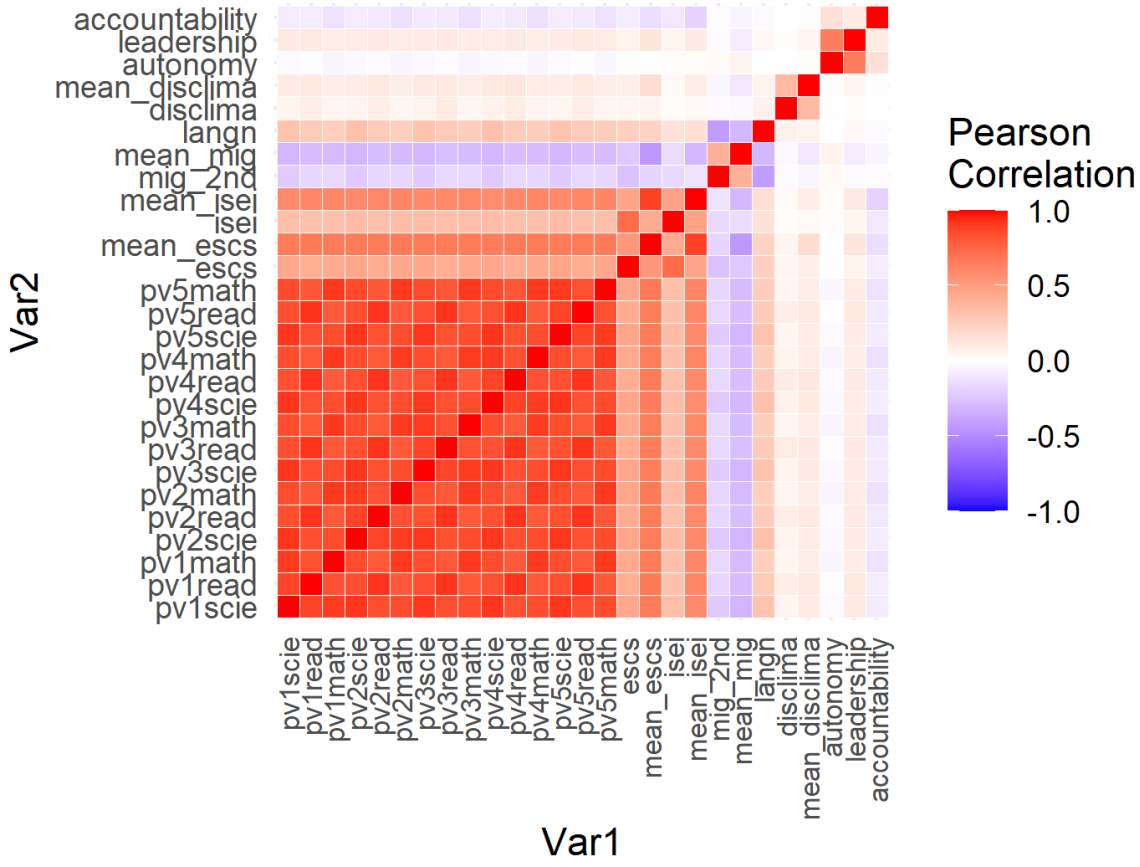


Figure B6: Heatmap of the correlations among independent and dependent variables used in the regression models for Germany in 2015. Own calculations.

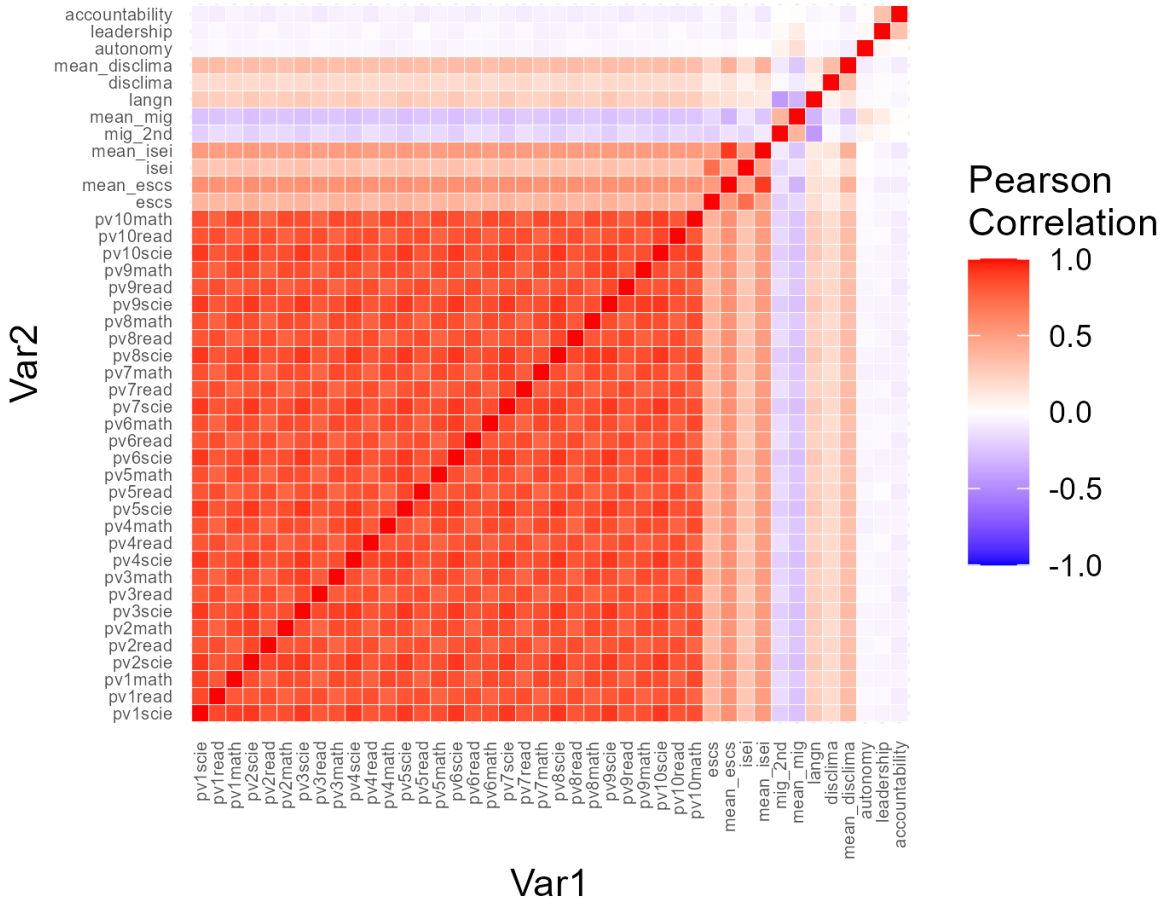


Figure B7: Heatmap of the correlations among independent and dependent variables used in the regression models for Sweden in 2000. Own calculations.

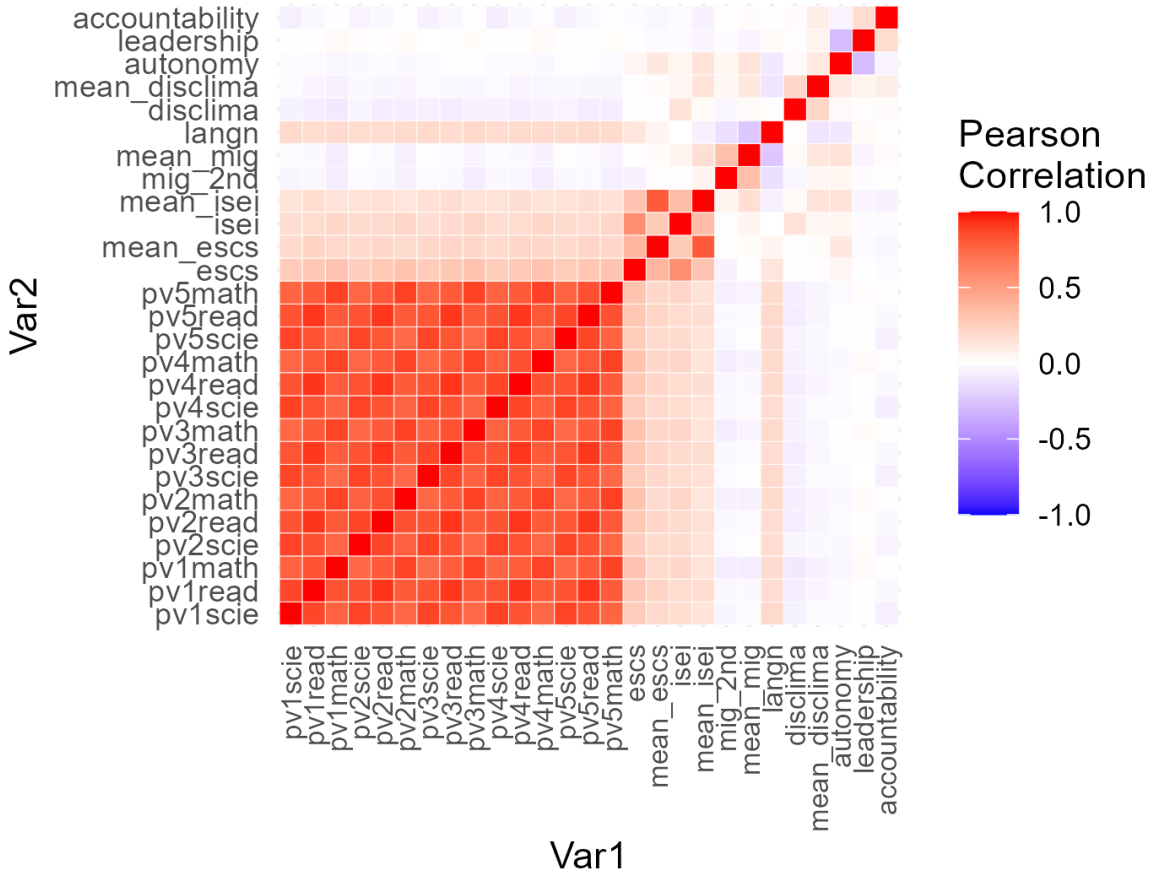


Figure B8: Heatmap of the correlations among independent and dependent variables used in the regression models for Sweden in 2009. Own calculations.

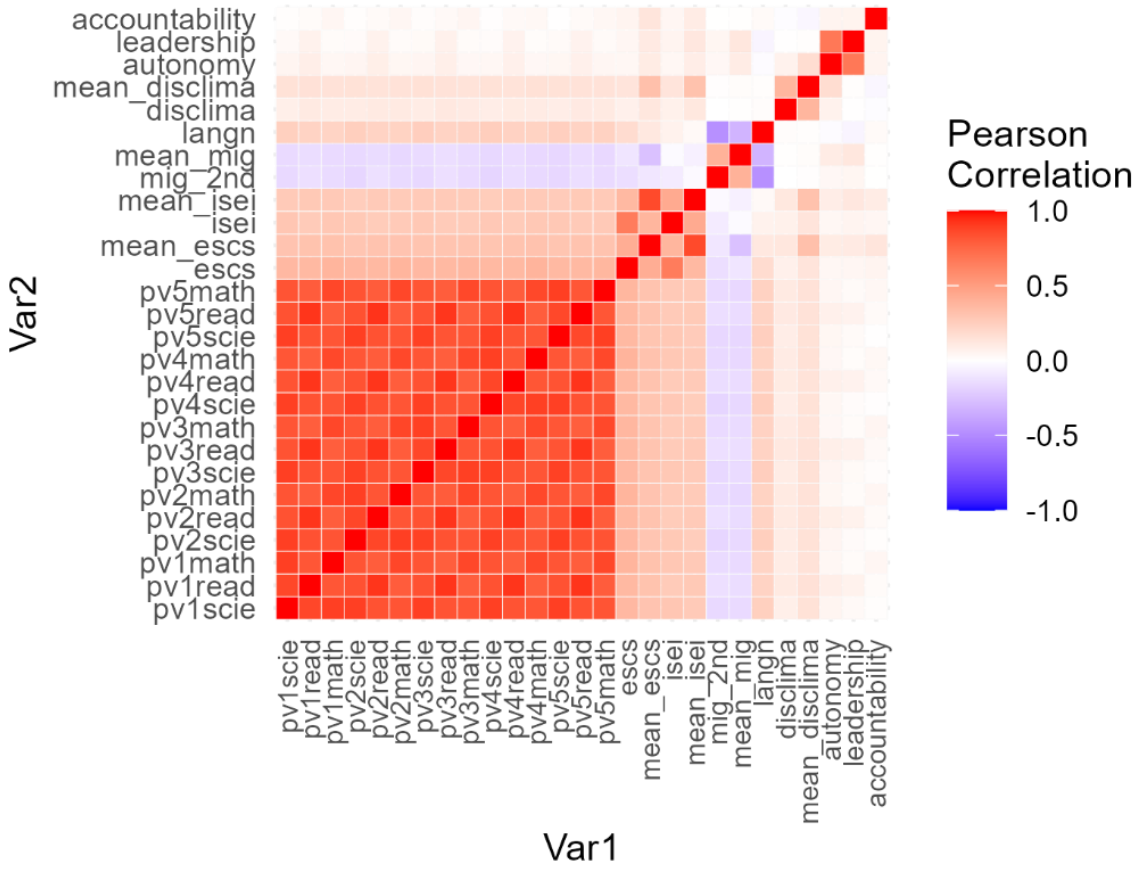


Figure B9: Heatmap of the correlations among independent and dependent variables used in the regression models for Sweden in 2015. Own calculations.

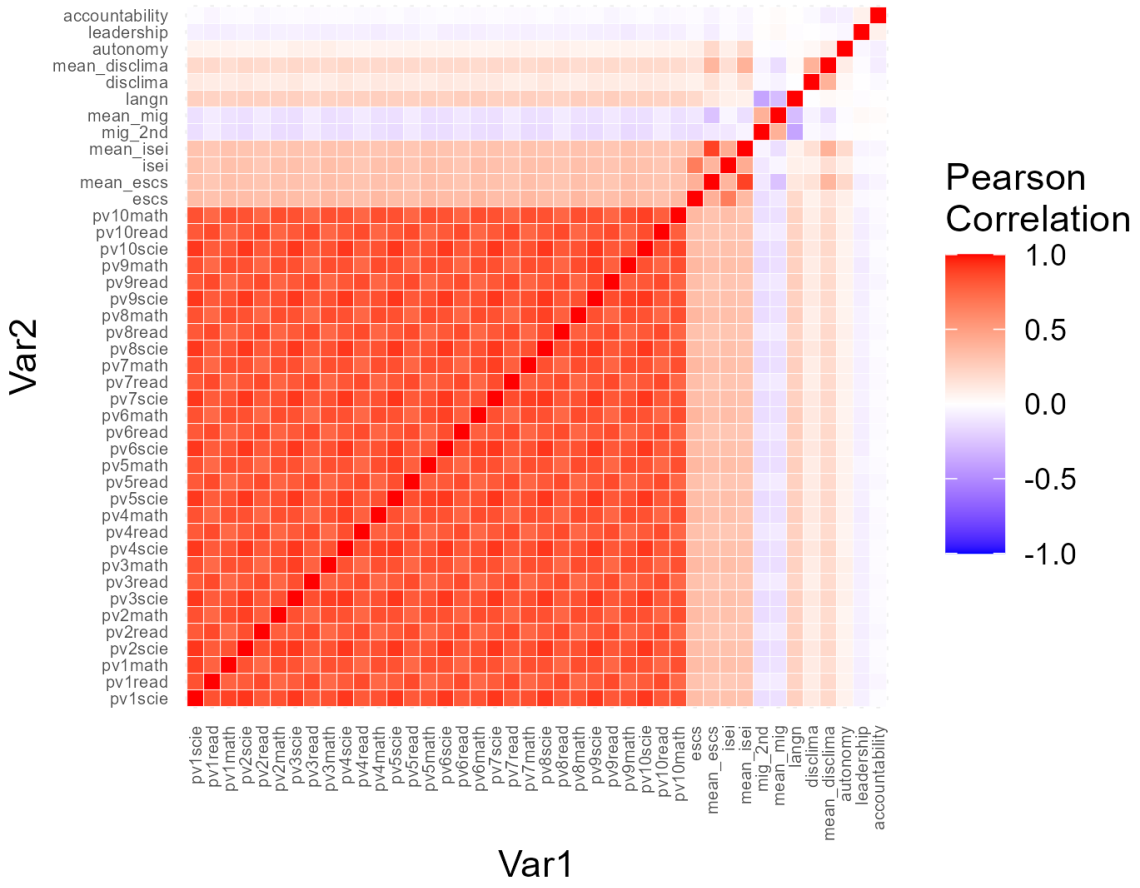


Figure B10: Heatmap of the correlations among independent and dependent variables used in the regression models for Finland in 2000. Own calculations.

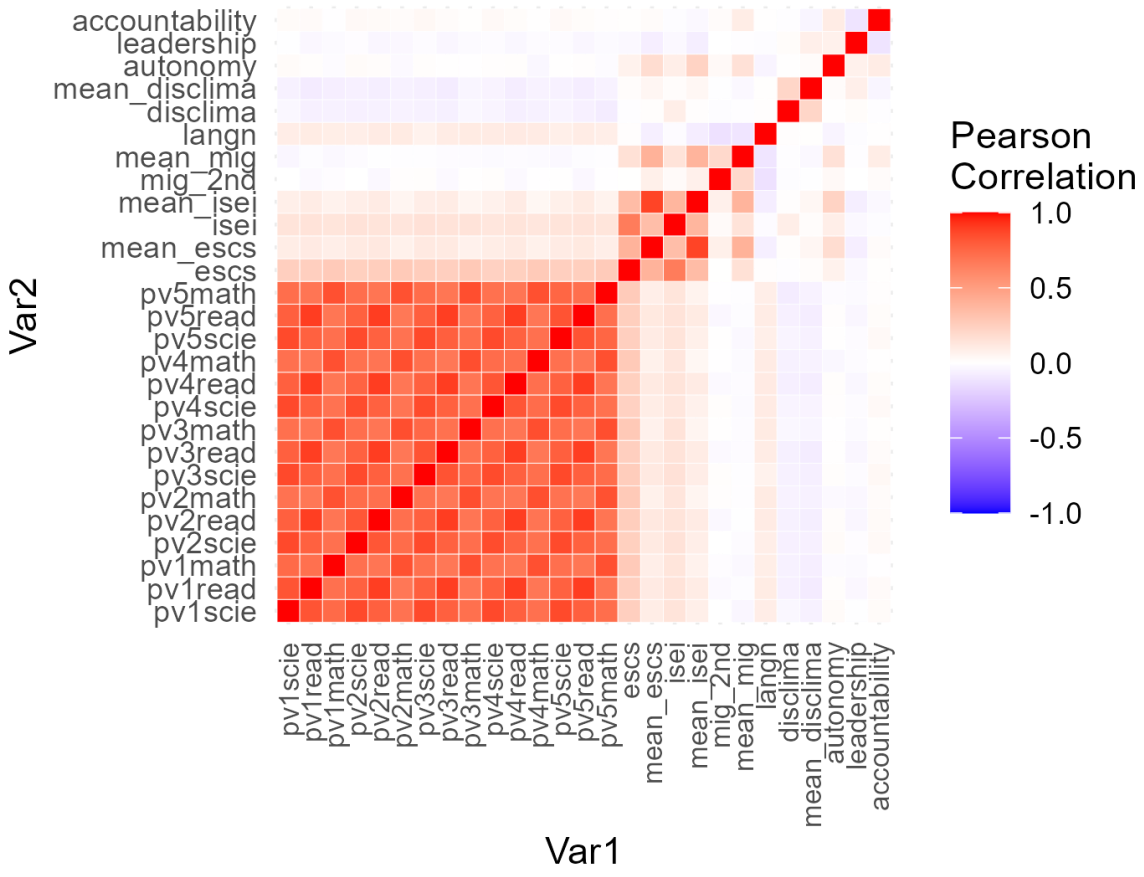


Figure B11: Heatmap of the correlations among independent and dependent variables used in the regression models for Finland in 2009. Own calculations.

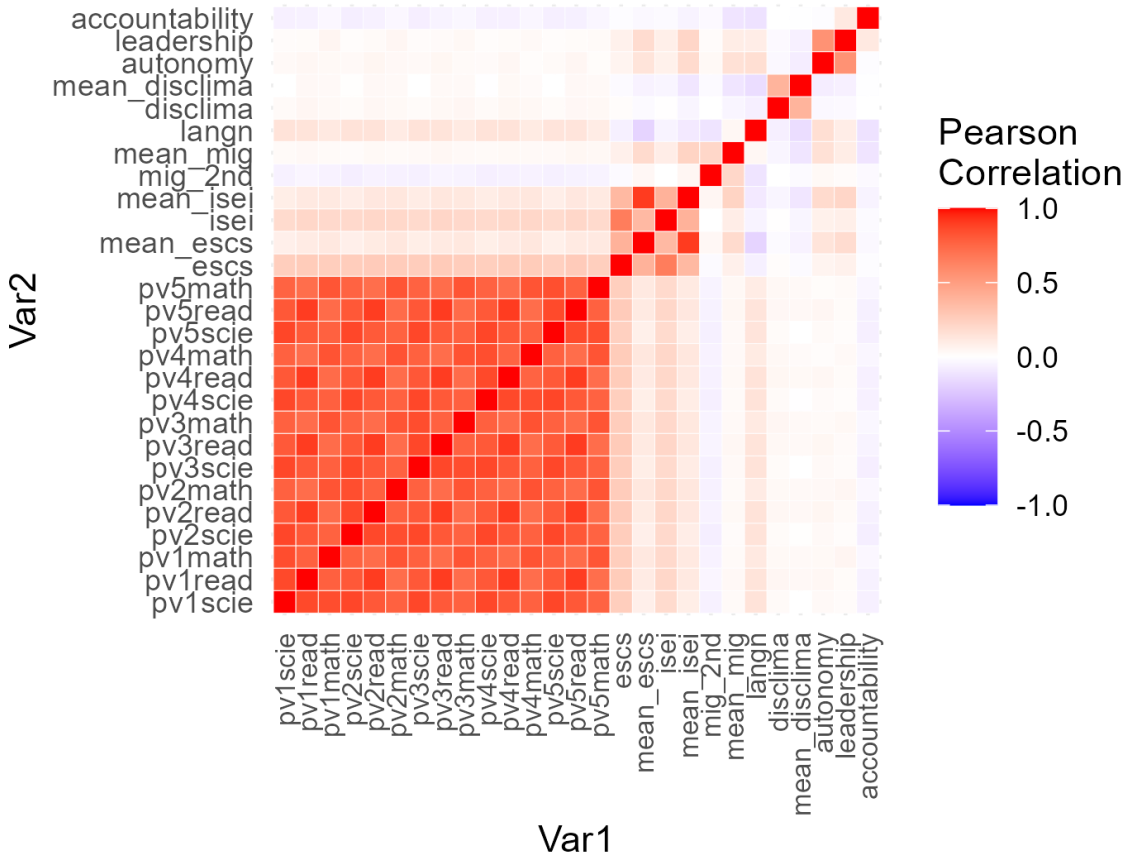
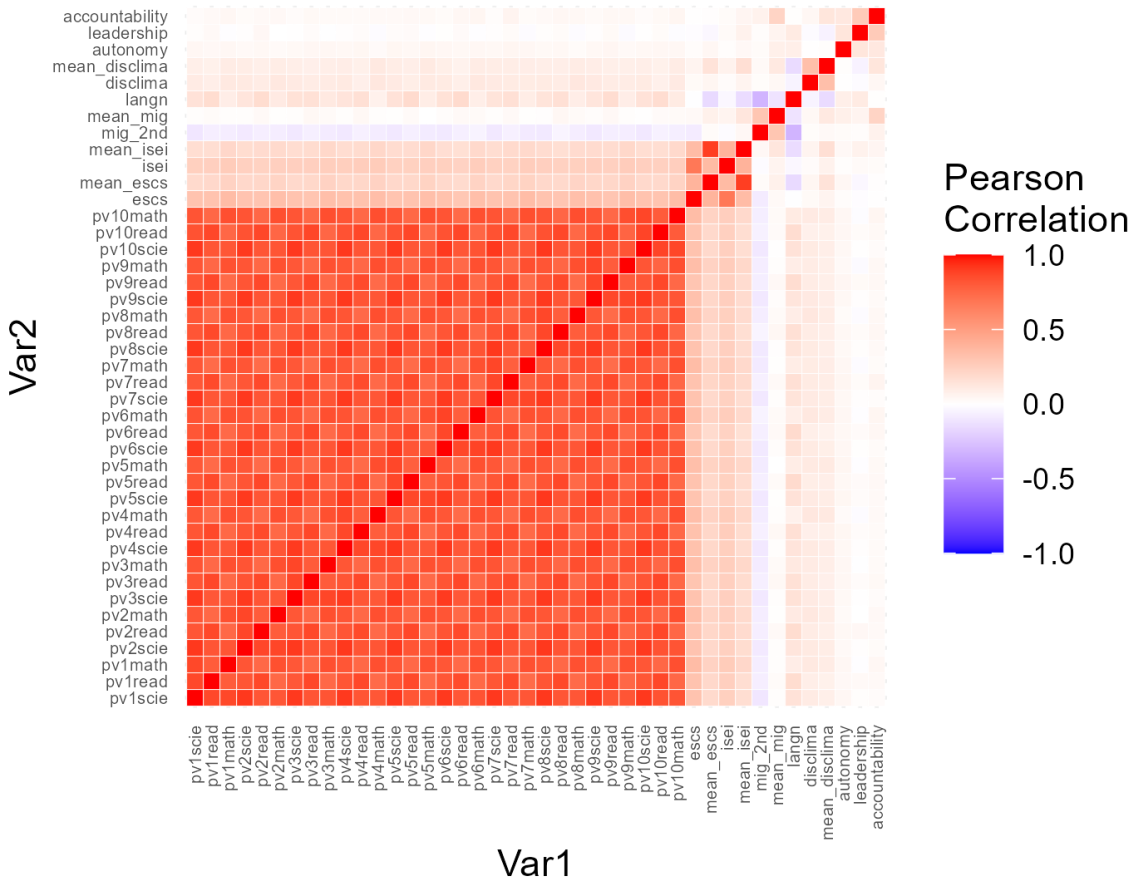


Figure B12: Heatmap of the correlations among independent and dependent variables used in the regression models for Finland in 2015. Own calculations.



C Regression Models

In a next step, we present the regression models calculated in tables C1 to C12. Note that due to imputations, every variable and goodness of fit measure yields significance values. If, for example, the R^2 values yield no significance, then we have to assume that these values are equal to zero and consequently that the model under scrutiny does not exhibit any variance explanatory power. As every ICC value calculated differs significantly from zero, we further assume that our multilevel models perform better compared to linear regression models. Tables C1-C3 report the results of the multilevel regression analysis for UK, C4-C6 for Germany, C7-C9 for Sweden, and finally, C10-C12 for Finland.

Table C1: Mean PISA score 2000, UK.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		32.124***		29.411***				27.868***	29.207***	27.763***
School level: Mean ESCS			92.782***	65.131***				64.937***	62.6***	62.924***
Student level ESCS * School level: Mean ESCS				3.638				2.845	2.879	2.463
Student level: Migrant background second generation					-3.371			-1.09		-1.078
School level: % Migrant background second generation					52.454			34.994*		30.619*
Language at home = Test language					34.735***			18.802**		19.242**
Student level: School disciplinary climate (WLE)						-2.34***		-1.994***		-1.983***
School level: Mean disciplinary climate						-1.529		0.732		0.776
School autonomy							15.903**		3.108	1.475
Educational leadership							-767.067*		-337.083*	-351.271*
Accountability							-11.311		-6.718*	-6.252*
Intercept	531.624***	530.94***	552.95***	545.337***	550.478***	533.72***	469.273***	561.591***	514.444***	524.034***
Students	5805	5805	5805	5805	5805	5805	5805	5805	5805	5805
Schools	232	232	232	232	232	232	232	232	232	232
ICC	0.288***	0.194***	0.093***	0.097***	0.274***	0.286***	0.27***	0.093***	0.09***	0.087***
Snijders-Bosker R2 level 1	0	0.144***	0***	0.116***	0.014	0.03	0	0.139***	0.116***	0.139***
Snijders-Bosker R2 level 2	0	0.116***	0.753***	0.753***	0.017	0.013	0.091	0.766***	0.772***	0.782***
R^2 Total	0	0.138***	0.22***	0.292***	0.014	0.025	0.026	0.311***	0.298***	0.315***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C2: Mean PISA score 2009, UK.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		30.092***		28.253***				26.403***	28.204***	26.365***
School level: Mean ESCS			100.088***	72.95***				74.261***	72.348***	73.35***
Student level ESCS * School level: Mean ESCS				5.167				5.545	5.002	5.412
Student level: Migrant background second generation					11.411			9.539		9.532
School level: % Migrant background second generation					-59.546			-17.561		-20.853
Language at home = Test language					24.382***			20.289***		20.216***
Student level: School disciplinary climate (WLE)						15.608***		14.247***		14.251***
School level: Mean disciplinary climate						19.839*		-17.192**		-17.268**
School autonomy							2.223		0.505	0.55
Educational leadership							-47.352		-6.261	-6.611
Accountability							-3.545		-0.613	-1.134
Intercept	502.101***	503.858***	532.743***	524.995***	459.177***	503.597***	303.643*	497.705***	496.907***	467.469***
Students	8480	8480	8480	8480	8480	8480	8480	8480	8480	8480
Schools	354	354	354	354	354	354	354	354	354	354
ICC	0.275***	0.2***	0.085***	0.093***	0.265***	0.264***	0.268***	0.092***	0.092***	0.092***
Snijders-Bosker R2 level 1	0	0.149***	0***	0.121***	0.014	0.083***	0	0.189***	0.121***	0.189***
Snijders-Bosker R2 level 2	0	0.079***	0.759***	0.75***	0.032	0.081	0.039	0.756***	0.751***	0.759***
R^2 Total	0	0.136***	0.212***	0.287***	0.019	0.082***	0.011	0.332***	0.288***	0.333***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C3: Mean PISA score 2015, UK.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		23.601***		22.192***				20.933***	20.933***	20.989***
School level: Mean ESCS			85.713***	65.648***				64.802***	65.355***	64.553***
Student level ESCS * School level: Mean ESCS				4.966				5.845	5.096	5.993
Student level: Migrant background second generation					-2.801			0.017		0.008
School level: % Migrant background second generation					11.261			-2.327		-4.428
Language at home = Test language					15.016*			11.43		11.4
Student level: School disciplinary climate (WLE)						15.94***		14.997***		14.992***
School level: Mean disciplinary climate						18.843		-3.471		-3.993
School autonomy							70.172		-23.767	-23.215
Educational leadership							-4.356		0.879	0.334
Accountability							-7.749**		-1.606	-1.257
Intercept	507.549***	509.844***	543.883***	536.334***	498.949***	506.124***	1212.562	524.871***	302.906	295.838
Students	7126	7126	7126	7126	7126	7126	7126	7126	7126	7126
Schools	320	320	320	320	320	320	320	320	320	320
ICC	0.252***	0.186***	0.078***	0.082***	0.252***	0.242***	0.235***	0.082***	0.081***	0.081***
Snijders-Bosker R^2 level 1	0	0.119***	0***	0.09***	0.007	0.096***	0	0.172***	0.09***	0.172***
Snijders-Bosker R^2 level 2	0	0.058***	0.756***	0.75***	0	0.073	0.214	0.757***	0.757***	0.764***
R^2 Total	0	0.109***	0.195***	0.252***	0.005	0.091***	0.06	0.308***	0.255***	0.312***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C4: Mean PISA score 2000, Germany.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		15.164***		11.571***				10.414***	11.512***	10.343***
School level: Mean ESCS			135.492***	120.756***				104.179***	124.549***	105.524***
Student level ESCS * School level: Mean ESCS				-5.625				-1.822	-5.719	-1.938
Student level: Migrant background second generation					-8.63*			-6.533*		-6.626*
School level: % Migrant background second generation					-3.396			-30.704		-26.143
Language at home = Test language					56.917***			47.962***		47.739***
Student level: School disciplinary climate (WLE)						-1.001***		-0.946***		-0.946***
School level: Mean disciplinary climate						-1.031**		0.024		0.019
School autonomy							-3.375		-6.749	-6.111
Educational leadership							-7.064		-210.649	-178.001
Accountability							-14.488*		2.825	-1.091
Intercept	482.415***	484.728***	542.567***	539.492***	427.378***	492.737***	484.046***	463.675***	489.806***	425.854***
Students	4178	4178	4178	4178	4178	4178	4178	4178	4178	4178
Schools	198	198	198	198	198	198	198	198	198	198
ICC	0.63***	0.597***	0.289***	0.293***	0.619***	0.553***	0.62***	0.255***	0.287***	0.25***
Snijders-Bosker R^2 level 1	0	0.043***	0***	0.038**	0.057**	0.005	0	0.079***	0.038**	0.079***
Snijders-Bosker R^2 level 2	0	0.008***	0.735***	0.734***	0.006	0.155***	0.033	0.767***	0.742***	0.775***
R^2 Total	0	0.023**	0.445***	0.455***	0.026*	0.094***	0.021	0.475***	0.461***	0.48***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C5: Mean PISA score 2009, Germany.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		15.898***		12.68***				10.246***	12.671***	10.23***
School level: Mean ESCS			130.167***	115.925***				118.579***	113.329***	116.308***
Student level ESCS * School level: Mean ESCS				-5.878				-5.185	-5.905	-5.232
Student level: Migrant background second generation					-14.502***			-9.106*		-9.081*
School level: % Migrant background second generation					-167.545**			28.818		31.761
Language at home = Test language					25.998***			22.621***		22.666***
Student level: School disciplinary climate (WLE)						3.639**		2.906**		2.913**
School level: Mean disciplinary climate						7.361		-13.981		-13.821
School autonomy							-3.559		-1.509	-1.599
Educational leadership							115.195*		36.067	37.146
Accountability							-8.358		-1.368	-1.191
Intercept	514.121***	515.553***	552.68***	550.596***	355.077***	517.458***	782.884***	550.657***	633.006***	637.888***
Students	3918	3918	3918	3918	3918	3918	3918	3918	3918	3918
Schools	191	191	191	191	191	191	191	191	191	191
ICC	0.577***	0.541***	0.264***	0.271***	0.539***	0.577***	0.561***	0.27***	0.266***	0.265***
Snijders-Bosker R^2 level 1	0	0.047***	0***	0.039***	0.04***	0.004	0	0.067***	0.039***	0.067***
Snijders-Bosker R^2 level 2	0	0.012***	0.748***	0.749***	0.143*	0.004	0.1	0.756***	0.755***	0.763***
R^2 Total	0	0.028***	0.44***	0.456***	0.098**	0.004	0.059	0.471***	0.46***	0.475***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C6: Mean PISA score 2015, Germany.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		15.556***		13.488***				10.585***	13.483***	10.589***
School level: Mean ESCS			110.349***	98.312***				89.284***	97.802***	88.976***
Student level ESCS * School level: Mean ESCS				3.111				1.087	3.066	1.091
Student level: Migrant background second generation					-15.348**			-10.859		-10.909
School level: % Migrant background second generation					-114.044			-29.1		-29.144
Language at home = Test language					41.577***			39.798***		39.731***
Student level: School disciplinary climate (WLE)						8.835***		8.014***		8.018***
School level: Mean disciplinary climate						63.45**		11.893		11.428
School autonomy							42.119		-7.715	2.229
Educational leadership							1.238		1.3	1.269
Accountability							-3.693		-2.002	-1.5
Intercept	497.601***	501.581***	539.218***	536.24***	361.564***	508.478***	727.001	475.846***	495.856**	488.737**
Students	2693	2693	2693	2693	2693	2693	2693	2693	2693	2693
Schools	148	148	148	148	148	148	148	148	148	148
ICC	0.528***	0.479***	0.196***	0.205***	0.491***	0.47***	0.526***	0.181***	0.203***	0.18***
Snijders-Bosker R^2 level 1	0	0.04**	0***	0.028**	0.05**	0.017**	0	0.082***	0.028**	0.082***
Snijders-Bosker R^2 level 2	0	0.015**	0.747***	0.744***	0.094	0.131	0.021	0.788***	0.748***	0.79***
R^2 Total	0	0.028**	0.367***	0.383***	0.072	0.074	0.011	0.427***	0.385***	0.428***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C7: Mean PISA score 2000, Sweden.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		30.726***		26.348***				24.02***	26.128***	23.74***
School level: Mean ESCS			65.692***	38.439***				36.988***	39.533***	37.554***
Student level ESCS * School level: Mean ESCS				-2.27				-2.128	-2.549	-2.481
Student level: Migrant background second generation					-5.494			-2.231		-2.237
School level: % Migrant background second generation					25.431			19.638		21.537
Language at home = Test language					64.274***			53.138***		53.008***
Student level: School disciplinary climate (WLE)						-0.795*		-0.775*		-0.777*
School level: Mean disciplinary climate						-1.394		-0.904		-0.728
School autonomy							0.292		-3.827	-2.201
Educational leadership							54.986		24.853	31.892
Accountability							-2.469		-2.371	-2.237
Intercept	515.512***	518.743***	567.707***	549.158***	485.389***	516.627***	524.345***	521.612***	553.97***	529.334***
Students	4192	4192	4192	4192	4192	4192	4192	4192	4192	4192
Schools	151	151	151	151	151	151	151	151	151	151
ICC	0.087***	0.056***	0.037**	0.042***	0.069***	0.087***	0.086***	0.029**	0.039***	0.028**
Snijders-Bosker R^2 level 1	0	0.078***	0***	0.067***	0.035***	0.002	0	0.092***	0.067***	0.092***
Snijders-Bosker R^2 level 2	0	0.157**	0.612***	0.592***	0.082	0.015	0.013	0.697***	0.617***	0.715***
R^2 Total	0	0.083***	0.056***	0.115***	0.038***	0.003	0.001	0.142***	0.117***	0.143***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C8: Mean PISA score 2009, Sweden.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		37.779***		35.387***				34.124***	35.407***	34.14***
School level: Mean ESCS			84.614***	50.221***				39.201***	50.12***	38.812***
Student level ESCS * School level: Mean ESCS				1.136				3.601	1.173	3.634
Student level: Migrant background second generation					-7.992			-3.765		-3.765
School level: % Migrant background second generation					-54.633			-24.169		-26.232
Language at home = Test language					54.901***			42.782***		42.807***
Student level: School disciplinary climate (WLE)						5.563**		4.61*		4.61*
School level: Mean disciplinary climate						40.712***		15.357*		14.278*
School autonomy							1.763		1.158	0.58
Educational leadership							3.532		-20.969	1.323
Accountability							1.386		-0.554	-0.478
Intercept	501.237***	503.855***	554.5***	535.31***	419.312***	497.203***	502.189**	473.671***	436.318***	474.865***
Students	4310	4310	4310	4310	4310	4310	4310	4310	4310	4310
Schools	183	183	183	183	183	183	183	183	183	183
ICC	0.16***	0.098***	0.049***	0.058***	0.147***	0.131***	0.156***	0.052***	0.057***	0.051***
Snijders-Bosker R^2 level 1	0	0.119***	0***	0.102***	0.045***	0.003	0	0.131***	0.102***	0.131***
Snijders-Bosker R^2 level 2	0	0.170***	0.717***	0.692***	0.131	0.196*	0.041	0.739***	0.700***	0.742***
R^2 Total	0	0.125***	0.11***	0.192***	0.059***	0.034*	0.007	0.225***	0.193***	0.225***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C9: Mean PISA score 2015, Sweden.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		34.296***		45.125***				40.247***	44.733***	39.973***
School level: Mean ESCS			89.461***	58.306***				49.825***	57.25***	48.972***
Student level ESCS * School level: Mean ESCS				18.253**				16.512**	17.759**	16.18**
Student level: Migrant background second generation					-16.74**			-9.966		-9.884
School level: % Migrant background second generation					27.564			39.58		36.229
Language at home = Test language					46.611***			36.873***		37.012***
Student level: School disciplinary climate (WLE)						4.466*		4.629**		4.632**
School level: Mean disciplinary climate						45.433***		18.385**		17.563**
School autonomy							-0.11		-21.258	-18.421
Educational leadership							-12.83**		-6.381*	-6.294*
Accountability							-2.59		-1.243	-0.64
Intercept	506.747***	509.266***	568.016***	547.063***	487.834***	507.345***	506.077*	538.641***	411.593**	418.613***
Students	4804	4804	4804	4804	4804	4804	4804	4804	4804	4804
Schools	192	192	192	192	192	192	192	192	192	192
ICC	0.168***	0.11***	0.053***	0.06***	0.173***	0.119***	0.158***	0.053***	0.057***	0.051***
Snijders-Bosker R^2 level 1	0	0.106***	0***	0.093***	0.05***	0.003	0	0.123***	0.092***	0.123***
Snijders-Bosker R^2 level 2	0	0.129***	0.713***	0.704***	0.013	0.3***	0.093	0.751***	0.721***	0.766***
R^2 Total	0	0.108***	0.116***	0.193***	0.044***	0.051**	0.016	0.227***	0.196***	0.23***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C10: Mean PISA score 2000, Finland.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		23.391***		22.638***				22.286***	22.654***	22.306***
School level: Mean ESCS			32.473**	10.616				17.385	11.975	18.584
Student level ESCS * School level: Mean ESCS				-4.016				-4.73	-3.694	-4.385
Student level: Migrant background second generation					3.169			8.222		8.121
School level: % Migrant background second generation					-1.177			-194.45*		-196.907*
Language at home = Test language					73.771***			69.959***		69.532***
Student level: School disciplinary climate (WLE)						-0.313		-0.283		-0.285
School level: Mean disciplinary climate						-1.312		-1.712*		-1.642
School autonomy							-3.313		-7.082	-5.741
Educational leadership							-46.008		-23.582	-2.932
Accountability							1.158		1.425	1.831
Intercept	541.749***	542.075***	543.371***	543.067***	468.137***	542.792***	533.157***	331.507***	540.759***	331.103***
Students	4421	4421	4421	4421	4421	4421	4421	4421	4421	4421
Schools	146	146	146	146	146	146	146	146	146	146
ICC	0.08*	0.065*	0.056*	0.063**	0.082*	0.077*	0.078*	0.055**	0.061**	0.054**
Snijders-Bosker R^2 level 1	0	0.076***	0	0.073***	0.013	0.001	0	0.086***	0.073***	0.086***
Snijders-Bosker R^2 level 2	0	0.136**	0.289*	0.264*	0.01	0.018	0.015	0.378***	0.299*	0.404***
R^2 Total	0	0.08***	0.022	0.088***	0.012	0.003	0.001	0.109***	0.091***	0.111***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C11: Mean PISA score 2009, Finland.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		29.219***		20.584**				19.638**	20.755**	19.864**
School level: Mean ESCS			40.679***	12.681				18.78**	12.498	19.201**
Student level ESCS * School level: Mean ESCS				-8.13				-8.35	-7.96	-8.121
Student level: Migrant background second generation					-30.271*			-23.223		-23.164
School level: % Migrant background second generation					82.731			-30.13		-30.061
Language at home = Test language					39.331***			40.402***		40.447***
Student level: School disciplinary climate (WLE)						2.329		1.831		1.829
School level: Mean disciplinary climate						13.963		15.545*		15.264*
School autonomy							0.153		-0.148	-0.186
Educational leadership							6.324		-0.643	-2.519
Accountability							-1.97		-1.337	-0.845
Intercept	543.544***	548.375***	582.487***	560.989***	547.499***	537.559***	560.951***	508.724***	561.861***	505.243***
Students	5672	5672	5672	5672	5672	5672	5672	5672	5672	5672
Schools	201	201	201	201	201	201	201	201	201	201
ICC	0.068***	0.058***	0.051***	0.058***	0.062***	0.065***	0.065***	0.043***	0.057***	0.042***
Snijders-Bosker R^2 level 1	0	0.091***	0*	0.089***	0.014*	0.001	0	0.101***	0.089***	0.101***
Snijders-Bosker R^2 level 2	0	0.132***	0.298**	0.263**	0.136**	0.073	0.031	0.456***	0.274**	0.462***
R^2 Total	0	0.094***	0.021*	0.101***	0.023**	0.006	0.002	0.126***	0.102***	0.126***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C12: Mean PISA score 2015, Finland.

	0	1	2	3	4	5	6	7	8	9
Student level: ESCS		36.595***		35.715***				32.677***	35.857***	32.903***
School level: Mean ESCS			67.583***	33.773***				39.203***	32.969***	38.787***
Student level ESCS * School level: Mean ESCS				2.271				-0.154	2.489	0.177
Student level: Migrant background second generation					-31.032**			-12.615		-13.027
School level: % Migrant background second generation					4.502			20.393		9.279
Language at home = Test language					45.633***			45.379***		44.745***
Student level: School disciplinary climate (WLE)						7.349***		7.1***		7.097***
School level: Mean disciplinary climate						1.785		7.757		7.424
School autonomy							30.758		19.287	7.925
Educational leadership							0.219		1.211	0.375
Accountability							1.7		1.131	0.998
Intercept	523.306***	528.459***	568.123***	549.821***	484.057***	522.735***	665.333***	518.953***	638.183***	550.691***
Students	4877	4877	4877	4877	4877	4877	4877	4877	4877	4877
Schools	144	144	144	144	144	144	144	144	144	144
ICC	0.091***	0.052***	0.036***	0.043***	0.107***	0.088**	0.088***	0.037***	0.041***	0.037***
Snijders-Bosker R^2 level 1	0	0.102***	0***	0.091***	0.023**	0.006	0	0.114***	0.091***	0.114***
Snijders-Bosker R^2 level 2	0	0.248***	0.636***	0.612***	0.097*	0.011	0.053	0.687***	0.633***	0.693***
R^2 Total	0	0.111***	0.059***	0.14***	0.031**	0.007	0.005	0.17***	0.142***	0.171***

Standardized beta coefficients
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D Regression Diagnostics

Technically, the multilevel linear regression models follow the same assumptions as linear regression models (Korendijk et al., 2008). These are linearity of the independent variables, normality of error terms, homoscedasticity, and independence of observations.

Beginning with the latter, we cannot guarantee independence of observations on student level. As the students are cluster sampled, meaning that only whole classes were included in the PISA test, we used a two-level model to account for the assumption of independency of observations (Maas & Hox, 2005, p. 87). As the tested students are part of the same class taught by the same teachers, students have an impact on each other and values of variables such as disciplinary climate in classes may be determined by test design. In other words, this approach allows the residuals of the students to be correlated with the values of the student level variables and with school level variables. This in turn allows us to calculate two error terms (one on student level and one on school level) and thus to explain the portion of variance explained by a regression model on student as well as on school level (Snijders und Bosker, 1994).

We used the variance inflation factor (VIF) to test for heteroscedasticity of our models. To accomplish this, we programmed an own function to calculate the VIF for the two-level models used by the BIFIEsurvey package called vif_calculator. The VIF regresses each independent variable of a model under investigation on the standard errors of a regression model containing all other variables. For example, in a model with PISA science scores as dependent variable, the student's ESCS, the average ESCS at school, and the disciplinary climate at school, the VIF would test if the student's ESCS could predict the residuals of a regression model consisting of the average ESCS and the disciplinary climate. This procedure is then repeated for every independent variable.

Mathematically, the VIF is calculated for each variable separately by dividing 1 by $1-R^2$ for each regression with one of the independent variables included as dependent variable, whereas the rest acts as independent variable in this regression. The VIF is minimal at a value of 1 (no multicollinearity) and gets larger the more variance of the error terms is explained by a single variable (O'brien 2007). A VIF under 4 indicates low levels of multicollinearity, values between 4 and 10 are associated with medium levels of multicollinearity, and values over 10 indicate high levels of multicollinearity. If at least one variable achieves values over 10, chances are high that confidence intervals of the regressors (β) are miscalculated. A miscalculation of confidence intervals in turn renders the significance level unreliable. We might therefore either miss significant associations between independent variables and the PISA mean score, or worse, might understand associations wrongly.

The VIF, however, is a measure for multicollinearity for single level linear regression models. Therefore, variables associated with the school level may show very high values or may be multicollinear on a single level but not on the two levels included in our calculations. The VIF-values are reported in tables D1 to D12 for each model separately. The findings are reported in the same order as the descriptive statistics and regression analyses in appendix B and C. Overall, the models performed well and showed a low degree of multicollinearity with values < 4, with the exception of model 3 for Sweden 2000, models 3 and 7 in the case of Finland 2009 and model 3 for Finland 2015.

Table D1: Variance inflation factor UK 2000.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.855				1.901	1.864	1.908
School level: Mean ESCS			1.377				1.418	1.455	1.495
Student level: ESCS * School level: Mean ESCS			1.454				1.573	1.482	1.596
Student level: Migrant background second generation				1.179			1.184		1.184
School level: % Migrant background second generation				1.226			1.310		1.324
Language at home = Test language				1.104			1.125		1.129
Student level: School disciplinary climate (WLE)					1.008		1.013		1.013
School level: Mean disciplinary climate					1.008		1.039		1.045
School autonomy						1.016		1.077	1.099
Educational leadership						1.008		1.020	1.024
Accountability						1.008		1.039	1.039
Mean VIF			1.562	1.170	1.008	1.011	1.320	1.323	1.260

Table D2: Variance inflation factor UK 2009.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.944				1.965	1.948	1.968
School level: Mean ESCS			1.305				1.476	1.317	1.496
Student level: ESCS * School level: Mean ESCS			1.601				1.632	1.611	1.639
Student level: Migrant background second generation				1.305			1.305		1.305
School level: % Migrant background second generation				1.327			1.349		1.387
Language at home = Test language				1.06			1.064		1.066
Student level: School disciplinary climate (WLE)					1.161		1.168		1.168
School level: Mean disciplinary climate					1.161		1.324		1.328
School autonomy						1.626		1.642	1.656
Educational leadership						1.645		1.646	1.653
Accountability						1.015		1.021	1.036
Mean VIF			1.617	1.231	1.161	1.429	1.41	1.531	1.427

Table D3: Variance inflation factor UK 2015.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			2.381				2.404	2.386	2.411
School level: Mean ESCS			1.29				1.361	1.341	1.419
Student level: ESCS * School level: Mean ESCS			2.025				2.059	2.034	2.071
Student level: Migrant background second generation				1.268			1.268		1.268
School level: % Migrant background second generation				1.268			1.296		1.316
Language at home = Test language				1.036			1.038		1.039
Student level: School disciplinary climate (WLE)					1.162		1.167		1.167
School level: Mean disciplinary climate					1.162		1.272		1.304
School autonomy						1.014		1.032	1.058
Educational leadership						1.031		1.055	1.065
Accountability						1.017		1.032	1.05
Mean VIF			1.899	1.19	1.162	1.021	1.483	1.48	1.379

Table D4: Variance inflation factor Germany 2000.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.793				1.838	1.795	1.841
School level: Mean ESCS			1.401				1.495	1.537	1.656
Student level: ESCS * School level: Mean ESCS			1.45				1.548	1.458	1.558
Student level: Migrant background second generation				1.096			1.105		1.106
School level: % Migrant background second generation				1.106			1.162		1.174
Language at home = Test language				1.062			1.126		1.129
Student level: School disciplinary climate (WLE)					3.448		3.455		3.458
School level: Mean disciplinary climate					3.448		3.553		3.553
School autonomy						1.095		1.1	1.106
Educational leadership						1.099		1.112	1.119
Accountability						1.005		1.129	1.141
Mean VIF			1.548	1.088	3.448	1.066	1.91	1.355	1.713

Table D5: Variance inflation factor Germany 2009.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.653				1.725	1.653	1.726
School level: Mean ESCS			1.347				1.622	1.41	1.718
Student level: ESCS * School level: Mean ESCS			1.332				1.356	1.334	1.36
Student level: Migrant background second generation				1.336			1.381		1.381
School level: % Migrant background second generation				1.189			1.434		1.491
Language at home = Test language				1.24			1.271		1.273
Student level: School disciplinary climate (WLE)					1.154		1.158		1.158
School level: Mean disciplinary climate					1.154		1.176		1.177
School autonomy						1.752		1.762	1.795
Educational leadership						1.732		1.778	1.787
Accountability						1.02		1.059	1.092
Mean VIF			1.444	1.255	1.154	1.501	1.39	1.499	1.451

Table D6: Variance inflation factor Germany 2015.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.517				1.571	1.517	1.572
School level: Mean ESCS			1.41				1.744	1.417	1.762
Student level: ESCS * School level: Mean ESCS			1.151				1.153	1.156	1.159
Student level: Migrant background second generation				1.342			1.372		1.372
School level: % Migrant background second generation				1.192			1.275		1.32
Language at home = Test language				1.27			1.28		1.282
Student level: School disciplinary climate (WLE)					1.13		1.135		1.135
School level: Mean disciplinary climate					1.13		1.454		1.457
School autonomy						1.003		1.005	1.046
Educational leadership						1.149		1.153	1.154
Accountability						1.146		1.157	1.159
Mean VIF			1.359	1.268	1.13	1.099	1.373	1.234	1.311

Table D7: Variance inflation factor Sweden 2000.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			8.003				8.022	8.033	8.054
School level: Mean ESCS			1.231				1.235	1.245	1.247
Student level: ESCS * School level: Mean ESCS			7.403				7.404	7.436	7.438
Student level: Migrant background second generation				1.113			1.118		1.118
School level: % Migrant background second generation				1.159			1.171		1.188
Language at home = Test language				1.066			1.098		1.104
Student level: School disciplinary climate (WLE)					1.006		1.008		1.008
School level: Mean disciplinary climate					1.006		1.027		1.052
School autonomy						1.095		1.109	1.155
Educational leadership						1.131		1.132	1.142
Accountability						1.036		1.042	1.048
Mean VIF			5.546	1.113	1.006	1.088	2.76	3.333	2.323

Table D8: Variance inflation factor Sweden 2009.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			4.416				4.53	4.425	4.534
School level: Mean ESCS			1.239				1.47	1.275	1.544
Student level: ESCS * School level: Mean ESCS			4.015				4.127	4.026	4.132
Student level: Migrant background second generation				1.41			1.418		1.419
School level: % Migrant background second generation				1.217			1.318		1.371
Language at home = Test language				1.325			1.351		1.353
Student level: School disciplinary climate (WLE)					1.166		1.167		1.167
School level: Mean disciplinary climate					1.166		1.31		1.376
School autonomy						1.796		1.798	1.886
Educational leadership						1.796		1.809	1.892
Accountability						1.004		1.023	1.035
Mean VIF			3.224	1.317	1.166	1.532	2.086	2.393	1.974

Table D9: Variance inflation factor Sweden 2015.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			6.474				6.582	6.477	6.585
School level: Mean ESCS			1.206				1.429	1.249	1.476
Student level: ESCS * School level: Mean ESCS			6.208				6.281	6.212	6.286
Student level: Migrant background second generation				1.329			1.34		1.341
School level: % Migrant background second generation				1.218			1.287		1.287
Language at home = Test language				1.206			1.243		1.243
Student level: School disciplinary climate (WLE)					1.192		1.193		1.193
School level: Mean disciplinary climate					1.192		1.378		1.384
School autonomy						1.005		1.041	1.041
Educational leadership						1.007		1.013	1.015
Accountability						1.009		1.01	1.014
Mean VIF			4.629	1.251	1.192	1.007	2.592	2.833	2.17

Table D10: Variance inflation factor Finland 2000.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			1.18				1.183	1.18	1.183
School level: Mean ESCS			1.199				1.309	1.227	1.332
Student level: ESCS * School level: Mean ESCS			1.027				1.031	1.03	1.034
Student level: Migrant background second generation				1.041			1.043		1.043
School level: % Migrant background second generation				1.043			1.153		1.174
Language at home = Test language				1.023			1.027		1.029
Student level: School disciplinary climate (WLE)					1.021		1.023		1.023
School level: Mean disciplinary climate					1.021		1.025		1.035
School autonomy						1.013		1.034	1.044
Educational leadership						1.025		1.034	1.042
Accountability						1.023		1.024	1.034
Mean VIF			1.135	1.036	1.021	1.02	1.099	1.088	1.089

Table D11: Variance inflation factor Finland 2009.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			10.183				10.276	10.189	10.287
School level: Mean ESCS			1.2				1.293	1.241	1.343
Student level: ESCS * School level: Mean ESCS			9.756				9.853	9.762	9.865
Student level: Migrant background second generation				1.061			1.064		1.065
School level: % Migrant background second generation				1.049			1.113		1.136
Language at home = Test language				1.019			1.092		1.155
Student level: School disciplinary climate (WLE)					1.181		1.182		1.182
School level: Mean disciplinary climate					1.181		1.22		1.223
School autonomy						1.463		1.467	1.517
Educational leadership						1.482		1.507	1.51
Accountability						1.023		1.026	1.057
Mean VIF			7.046	1.043	1.181	1.322	3.387	4.199	2.849

Table D12: Variance inflation factor Finland 2015.

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8	Model9
Student level: ESCS			6.084				6.135	6.096	6.15
School level: Mean ESCS			1.192				1.242	1.2	1.255
Student level: ESCS * School level: Mean ESCS			5.795				5.82	5.81	5.839
Student level: Migrant background second generation				1.214			1.23		1.235
School level: % Migrant background second generation				1.107			1.118		1.195
Language at home = Test language				1.113			1.173		1.221
Student level: School disciplinary climate (WLE)					1.127		1.127		1.127
School level: Mean disciplinary climate					1.127		1.19		1.214
School autonomy						1.03		1.039	1.093
Educational leadership						1.096		1.096	1.114
Accountability						1.089		1.091	1.157
Mean VIF			4.357	1.144	1.127	1.072	2.379	2.722	2.055

Finally, we conducted Shapiro-Francia W' tests to account for the normality of the error terms assumption (Royston 1993; Thode Jr. 2002). We have to make sure that the normality of error terms of our models holds, as nonnormally distributed residuals (e.g., bimodal, multimodal) yield the danger of inaccurate prediction intervals (standard errors) of the independent variables (β). Therefore, our models might lead to miscalculated significance values. Additionally, the association between our dependent variable (PISA score) and our independent variables might not be linear.

As the BIFIE-survey package does not provide a prediction function and to extract error terms, we coded an own function using the following routine: First extract the beta values included in the models. Second, predict the values for each student based on the predictor values. Third, subtract the predicted values from each plausible value which are the basis of our regression models (e.g., the first mean plausible PISA score, the second etc.). Fourth, calculate the mean deviation from all given plausible values. At last, check if the number of observations is below or equal 5000. As the Shapiro-Francia W' test is specified for up to 5000 observations, we decided on drawing a sample of 5000 residuals as calculated above from our population if the number of observations exceeds this limit. Otherwise, the Shapiro-Francia W' test would yield biased estimates.

The null hypothesis (H_0) of the Shapiro-Francia W' test assumes that the residuals of a variable under consideration follow a normal distribution, whereas the alternate hypothesis (H_A) assumes that the residuals do not follow a normal distribution. Higher values of test statistics W and V indicate a violation of the normality assumption. Mathematically, the V value is like z-statistics in hypothesis testing. Therefore, the higher the V, the lower the p-value[1], as reported in the last column of tables D7 to D12, the more likely the residuals are not normally distributed. If a value of $p < 0.05$ is reported, the probability to reject H_0 albeit H_0 is true is less than 5% and the residuals of the respective variable follow most likely not a normal distribution.

As seen in tables D13 to D24 in the p-value column, the normality assumption is violated in most cases. This is possibly caused by our large sample sizes, as the Shapiro-Francia W' test tends to be more sensitive to violations of the normality assumption when the sample size increases. However, we decided to account for this possible violation by applying a heteroscedastic corrected covariance matrix estimation (Pek et al., 2018), also known as Huber-Sandwich estimator or robust standard errors (Freedman, 2006). Robust standard errors use the squared residual values for each observation instead of the constant and independent error terms usually applied in linear regression models. By doing so, we widen the confidence interval of the regressions considerably and allow variables to become insignificant more easily and to account for nonnormally distributed independent variables.

.

[1] Please note that the E in the p-value column translates to "exponent". For example, 3.57E-12 denotes $3.57*10^{-12}$.

Table D13: Shapiro-Francia test statistic for UK 2000. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.994875	3.57E-12	5	5805	yes	no
2	0.994684	1.93E-11	5	5805	yes	no
3	0.994062	2.46E-13	5	5805	yes	no
4	0.994745	2.29E-12	5	5805	yes	no
5	0.994919	4.16E-12	5	5805	yes	no
6	0.99527	1.45E-11	5	5805	yes	no
7	0.993115	1.45E-14	5	5805	yes	no
8	0.994022	2.17E-13	5	5805	yes	no
9	0.993669	7.36E-14	5	5805	yes	no

Table D14: Shapiro-Francia test statistic for UK 2009. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.996707	4.9E-09	5	8480	yes	no
2	0.998128	5.42E-07	5	8480	yes	no
3	0.998164	1.05E-05	5	8480	yes	no
4	0.998118	7.91E-06	5	8480	yes	no
5	0.99799	3.67E-06	5	8480	yes	no
6	0.998237	1.65E-05	5	8480	yes	no
7	0.998894	0.001484	5	8480	yes	no
8	0.998276	2.12E-05	5	8480	yes	no
9	0.998899	0.001539	5	8480	yes	no

Table D15: Shapiro-Francia test statistic for UK 2015. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.997998	3.84E-06	10	7126	yes	no
2	0.99699956	1.915E-08	10	7126	yes	no
3	0.998199	1.3E-05	10	7126	yes	no
4	0.996972	1.68E-08	10	7126	yes	no
5	0.997607	4.22E-07	10	7126	yes	no
6	0.997646	5.2E-07	10	7126	yes	no
7	0.998166	1.06E-05	10	7126	yes	no
8	0.998622	0.000209	10	7126	yes	no
9	0.998421	5.38E-05	10	7126	yes	no

Table D16: Shapiro-Francia test statistic for Germany 2000. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.991698	1.85E-14	5	4178	no	no
2	0.991698	1.85E-14	5	4178	no	no
3	0.99625	1.2E-08	5	4178	no	no
4	0.991698	1.85E-14	5	4178	no	no
5	0.991698	1.85E-14	5	4178	no	no
6	0.991698	1.85E-14	5	4178	no	no
7	0.997145	4.88E-07	5	4178	no	no
8	0.996859	1.4E-07	5	4178	no	no
9	0.997498	2.48E-06	5	4178	no	no

Table D17: Shapiro-Francia test statistic for Germany 2009. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.991866	1.05E-13	5	3918	no	no
2	0.991866	1.05E-13	5	3918	no	no
3	0.996567	1.06E-07	5	3918	no	no
4	0.991866	1.05E-13	5	3918	no	no
5	0.991866	1.05E-13	5	3918	no	no
6	0.991866	1.05E-13	5	3918	no	no
7	0.997317	2.39E-06	5	3918	no	no
8	0.99634	4.42E-08	5	3918	no	no
9	0.997029	6.89E-07	5	3918	no	no

Table D18: Shapiro-Francia test statistic for Germany 2015. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.990993	2.34E-11	10	2693	no	no
2	0.990993	2.34E-11	10	2693	no	no
3	0.996722	1.93E-05	10	2693	no	no
4	0.990993	2.34E-11	10	2693	no	no
5	0.990993	2.34E-11	10	2693	no	no
6	0.990993	2.34E-11	10	2693	no	no
7	0.998108	0.002755	10	2693	no	no
8	0.996932	3.85E-05	10	2693	no	no
9	0.998236	0.004606	10	2693	no	no

Table D19: Shapiro-Francia test statistic for Sweden 2000. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.99305	4.84E-13	5	4192	no	no
2	0.99305	4.84E-13	5	4192	no	no
3	0.99445	2.5E-11	5	4192	no	no
4	0.99305	4.84E-13	5	4192	no	no
5	0.99305	4.84E-13	5	4192	no	no
6	0.99305	4.84E-13	5	4192	no	no
7	0.995166	2.41E-10	5	4192	no	no
8	0.994528	3.16E-11	5	4192	no	no
9	0.995189	2.59E-10	5	4192	no	no

Table D20: Shapiro-Francia test statistic for Sweden 2009. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.997201	4.15E-07	5	4310	no	no
2	0.997201	4.15E-07	5	4310	no	no
3	0.998169	5.59E-05	5	4310	no	no
4	0.997201	4.15E-07	5	4310	no	no
5	0.997201	4.15E-07	5	4310	no	no
6	0.997201	4.15E-07	5	4310	no	no
7	0.998374	0.000182	5	4310	no	no
8	0.998254	9.05E-05	5	4310	no	no
9	0.99835	0.000158	5	4310	no	no

Table D21: Shapiro-Francia test statistic for Sweden 2015. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.995729	1.66E-10	10	4804	no	no
2	0.995729	1.66E-10	10	4804	no	no
3	0.995845	2.58E-10	10	4804	no	no
4	0.995729	1.66E-10	10	4804	no	no
5	0.995729	1.66E-10	10	4804	no	no
6	0.995729	1.66E-10	10	4804	no	no
7	0.997063	4.71E-08	10	4804	no	no
8	0.995895	3.13E-10	10	4804	no	no
9	0.997002	3.54E-08	10	4804	no	no

Table D21: Shapiro-Francia test statistic for Finland 2000. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.992978	1.35E-13	5	4421	no	no
2	0.992978	1.35E-13	5	4421	no	no
3	0.991691	5.33E-15	5	4421	no	no
4	0.992978	1.35E-13	5	4421	no	no
5	0.992978	1.35E-13	5	4421	no	no
6	0.992978	1.35E-13	5	4421	no	no
7	0.992852	9.69E-14	5	4421	no	no
8	0.992188	1.78E-14	5	4421	no	no
9	0.993256	2.87E-13	5	4421	no	no

Table D22: Shapiro-Francia test statistic for Finland 2009. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.992631	3.79E-15	5	5672	yes	no
2	0.992802	6.05E-15	5	5672	yes	no
3	0.992574	3.24E-15	5	5672	yes	no
4	0.993634	6.62E-14	5	5672	yes	no
5	0.992945	9E-15	5	5672	yes	no
6	0.992715	4.76E-15	5	5672	yes	no
7	0.99298	9.94E-15	5	5672	yes	no
8	0.992231	1.3E-15	5	5672	yes	no
9	0.993009	1.08E-14	5	5672	yes	no

Table D24: Shapiro-Francia test statistic for Finland 2015. Own calculations.

Model	W	P	imputations	obs	sampling	normally_distributed
1	0.994346	1E-12	10	4877	no	no
2	0.994346	1E-12	10	4877	no	no
3	0.994718	3.39E-12	10	4877	no	no
4	0.994346	1E-12	10	4877	no	no
5	0.994346	1E-12	10	4877	no	no
6	0.994346	1E-12	10	4877	no	no
7	0.996539	3.46E-09	10	4877	no	no
8	0.994961	7.75E-12	10	4877	no	no
9	0.99664	5.39E-09	10	4877	no	no

References

Avvisati, F. (2020). The measure of socio-economic status in PISA: a review and some suggested improvements. *Large-Scale Assessments in Education*, *8*(8), 1–37.

Freedman, D. A. (2006). On the so-called “Huber sandwich estimator” and “robust standard

errors". *The American Statistician*, *60*(4), 299–302.

Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, *6*(3), 223-239.

Korendijk, E. J. H., Maas, C. J. M., Moerbeek, M., & Van der Heijden, P. G. M. (2008). The influence of misspecification of the heteroscedasticity on multilevel regression parameter and standard error estimates. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*(2), 67-72.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6.

OECD. (2003). *PISA 2000 Technical Report*. Paris: OECD.

OECD. (2012). *PISA 2009 Technical Report*. Paris: OECD.

OECD. (2016). *PISA 2015 Results.* Vol. II. *Policies and Practices for Successful Schools*. Paris: OECD.

OECD. (2017). *PISA 2015 Technical Report*. Paris: OECD.

Pek, J., Wong, O., & Wong, A. C. M. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in Psychology*, *9*(November), 2104. https://doi.org/10.3389/fpsyg.2018.02104.

Robitsch, A., & Oberwimmer, K. (2019). BIFIEsurcey: Tools for survey statistics in educational assessment. R package version 3.3-12. https://CRAN.R-project.org/package=BIFIEsurvey, last retrieved on March 28, 2022.

Royston, P. (1993). A pocket-calculator algorithm for the Shapiro-Francia test for non-normality: an application to medicine. *Statistics in Medicine*, 12, 181-184.

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research*, *22*(3), 342–363.

Thode Jr., H. C. (2002). *Testing for Normality*. New York: Marcel Dekker.