The Social Geography of Education: Neighborhood, Class Composition, and the Educational Achievement of Elementary Students in Zurich, Switzerland

Online Appendix

Christoph Zangger

University of Bern, Department of Sociology of Education, Fabrikstrasse 8, CH-3012 Bern, Switzerland; christoph.zangger@edu.unibe.ch

Contents

I. Introduction

Studies relying on complete case analysis (i.e., if all cases with any missing values are omitted from the estimation) run the risk of drawing inference on a selective subgroup and of yielding biased estimates since the missing values might not be completely at random (MCAR – McKnight et al. 2007: 142f.). Hence, to ensure valid inference and additional statistical power, the missing pattern in the data needs to be adequately addressed. In what follows, the missing pattern, the imputation strategy, and the complete case analysis are briefly outlined.[1]

II. Missing Pattern

To start with, our main concern is to determine which missing values to impute. Table A1 gives a quick overview of the sample attrition in the two waves in the respective cities (unit-nonresponse). In Zurich, only about 53% of all parents responded in the first wave, their number dropping further in the second wave. However, as social background characteristics regarding parents' education and occupation were inquired only in the first wave, we decided to impute the missing values of item-nonresponse for the 525 cases in the first wave as it can be assumed that unit-nonresponse does not occur randomly and further assumptions about the missing pattern would be needed to adequately model these non-random missing values (Allison 2001; Proner 2011).

Tabelle A1: Sample Attrition

	N	%
Students in Wave 1	1541	100.00
Bern	555	36.02
Zürich	986	63.98
Students in Wave 2	1380	100.00
Bern	415	30.07
Zürich	965	69.93
Parents in Wave 1	846	100.00
Bern	321	37.94
Zürich	525	62.06
Parents in Wave 2	678	100.00
Bern	238	35.10
Zürich	440	64.90

[1] Throughout the documentation I will use the terms "estimation" or "analysis model" to denote the statistical model of the main paper and "imputation model" for the model, outlined in this document, aiming to adequately impute the missing values. Similarly, I will use the term "complete case analysis" to specify the statistical analysis based on a listwise deletion of all units with missing values, whereas "completed cases" will refer to the additional inclusion of the imputed values in the analysis.

Deleting cases with missing values in any of the included variables results in a reduction of another 172 cases (see Table A3 in the next section for descriptives of both imputed and listwise deleted cases). So the main reduction is caused by missing values in the dependent variable and the factor measuring students' problem solving skills (90 and 86 cases, respectively – Table A2). Hence, almost 38% of all cases in the first wave have one or several missing values in the model variables. About 11% of all cases have missing values in both, the dependent variable and the factor measuring problem solving skills. Another 5% of all cases have missing values only in the scale measuring the perceived deviance in the school context and about 4% have missing values in the social background characteristic, measured

Table A2: Missing Structure

Variables	Missing	Observed	Min.	Max.
Math Grade 6th School Year	90	435	2.5	6
Math Grade 5th School Year	14	511	2	6
Mean Grade in Math	13	512	3.77	5.15
Class Position (EGP)	32	493	1	5
Household's Financial Situation	6	519	1	3
Language prior to Enrolment	6	519	0	2
Problem Solving Skills	86	439	-3.099182	1.84721
Social Integration	15	510	-4.491384	1.142558
Perceived Deviance	38	487	-1.872782	3.852082
Gender	6	519	0	1

by parents' class position. Finally, 2% of all cases have missing values regarding the factor measuring integration into the local context and another 2% show a somewhat more complex pattern of missing values in the dependent variable, problem solving skills and average mathematical achievement. The remaining missing patterns each account for less than 1% of all cases. As can be seen from this brief description, the main item non-response does not point to a strongly selective missing pattern (e.g., parents of children with low grades also tend not to report their social background characteristics). Therefore, it seems justified to assume that – given observed covariates – the missing values are at random (MAR - Rubin 1987; Allison 2001; McKnight et al. 2007). Conditional on observed characteristics $X_{P\neq p}$, the probability of missing values in a variable $X_{P=p}$ is unrelated to the values of $X_{P=p}$ ($\Pr(X_{P=p\ missing}|X_{P=p}, X_{P\neq p}) = \Pr(X_{P=p\ missing}|X_{P\neq p})$ – Rubin 1987; Allison 2001). As Allison (2001: 5) notes, this condition is usually identical with the statement that the data generating process of the missing pattern is ignorable.

As a consequence, we are able to impute the missing values in the independent and in the dependent variable as a function of the observed values and additional fully observed covariates. The missing pattern, however, appears to be arbitrary and non-monotone (i.e., the missing values occur in a non-nested fashion; see Carpenter & Kenward 2013: 7; StataCorp 2013: 7). Hence, additional computational power will be needed to perform a multiple imputation of missing values – a procedure that will be elaborated in the next section.

III. Multiple Imputation

As can be seen in Table A2, there are ten variables in the statistical model exhibiting missing values. Hence, we need an imputation model which accounts for these missing patterns simultaneously and that adequately reproduces the underlying relationships between the model variables (Allison 2001; von Hippel 2009). For the very same reason, we need to include and impute the dependent variable; failing to do so would result in a different relationship between dependent and independent variables (Allison 2001: 35; Carpenter & Kenward 2013: 24ff.). Furthermore, all transformations and interaction terms of variables in the analysis model need to be included in the imputation equation (i.e., in the present case the interaction of the neighborhood characteristics with the scales measuring social integration and perceived deviance as well as with parent's class position – von Hippel 2009; Carpenter & Kenward 2013: 127ff.). Design variables, such as weights or clusters (i.e., in the present case the neighborhood cluster variable) should also form part of the imputation model (StataCorp 2013: 116). Finally, all fully observed variables of the estimation model as well as additional covariates that are assumed to be related to the missing pattern need to be included (Allison 2001: 35; StataCorp 2013: 8).

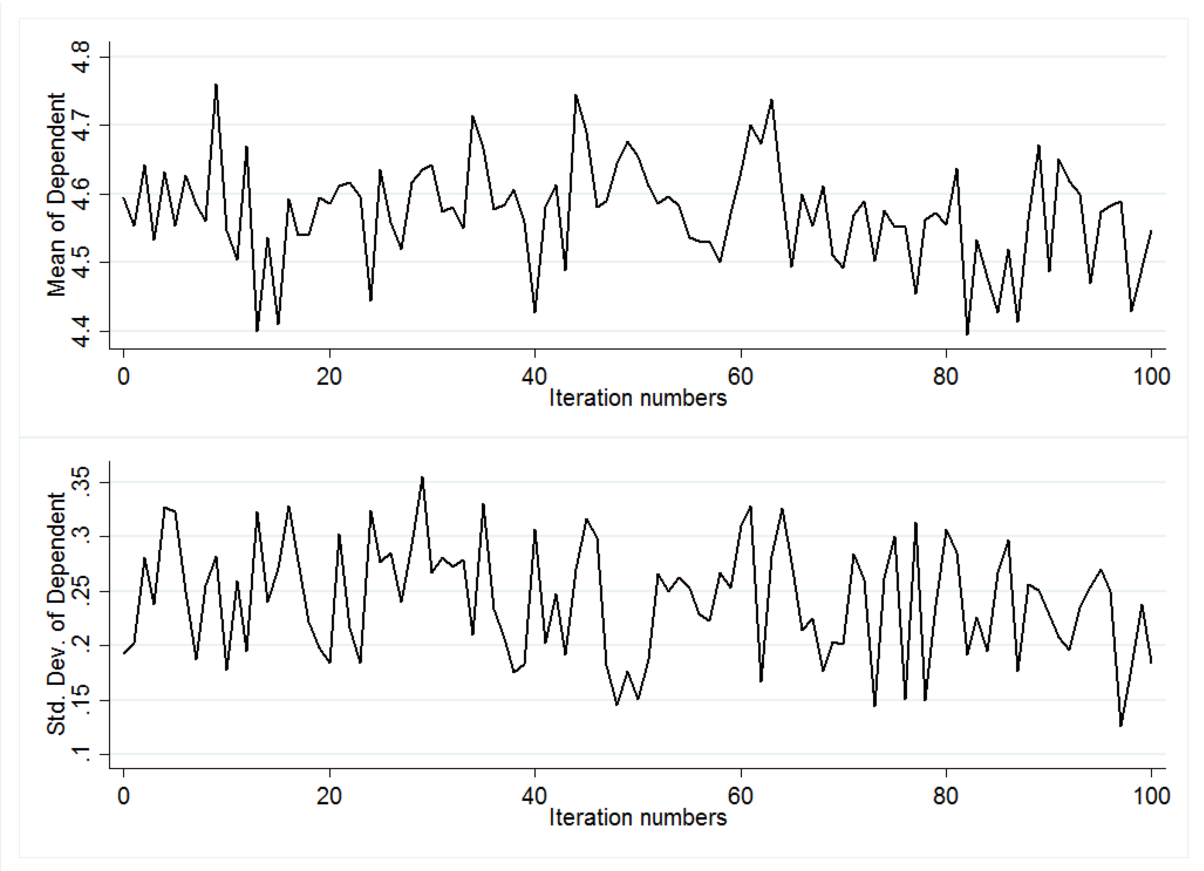
To impute the missing values in each variable, we use series of chained equations with additional, fully observed covariates (MICE – White et al. 2011; StataCorp 2013). Equation 1 describes the basic structure of the imputation model. Each variable with missing values

$$\begin{gathered} X_1^{t+1} \sim g_1(X_1|X_2^t, \dots, X_p^t, Z, \theta_1) \\ X_2^{t+1} \sim g_2(X_2|X_1^{t+1}, X_3^t, \dots, X_p^t, Z, \theta_2) \\ \dots \\ X_p^{t+1} \sim g_p(X_p|X_1^{t+1}, X_2^{t+1}, \dots, X_{p-1}^{t+1}, Z, \theta_p) \end{gathered} \tag{1}$$

$X_{P=p}$ is imputed as a function $g_p(\cdot)$ of all other model variables with missing values $X_{P\neq p}$ and the complete predictors Z, where θ_p denotes the model parameters with a uniform prior. Imputed values are drawn from this equation for iterations $t = 0,1, \dots, T$. The univariate imputation models $g_p(\cdot)$ can be of a different type (e.g., a multinomial logistic for categorical X_p with more than two categories). However, MICE requires that the conditional densities $g_1, g_2, \dots, g_p$ correspond to a multivariate joint conditional distribution of $X_1, X_2, \dots X_p$ given Z (StataCorp 2013: 143). In light of the many categorical variables with missing values, the use of MICE as a model with less prerequisites instead of a multivariate normal imputation model seems justified as the latter requires all variables to be normally distributed $X_p \sim N(\bar{x}_p, s_p)$ and linearly dependent on each other (Allison 2001: 33).

As computational power is no longer a scarce commodity, 200 imputed data sets were constructed in order to ensure reliable values. Within each imputation, 50 iterations were performed. The 50 iterations (burn in) were chosen as a conservative value to guarantee independence of each imputed value and to ensure convergence. As illustrated in Figure A1 for the dependent variable (plots for the other imputed variables are available from the author), there is no visible trend after about 20 to 30 iterations in neither the mean nor the standard deviation of the completed variable (Allison 2001: 43f.; StataCorp 2013: 153ff.).

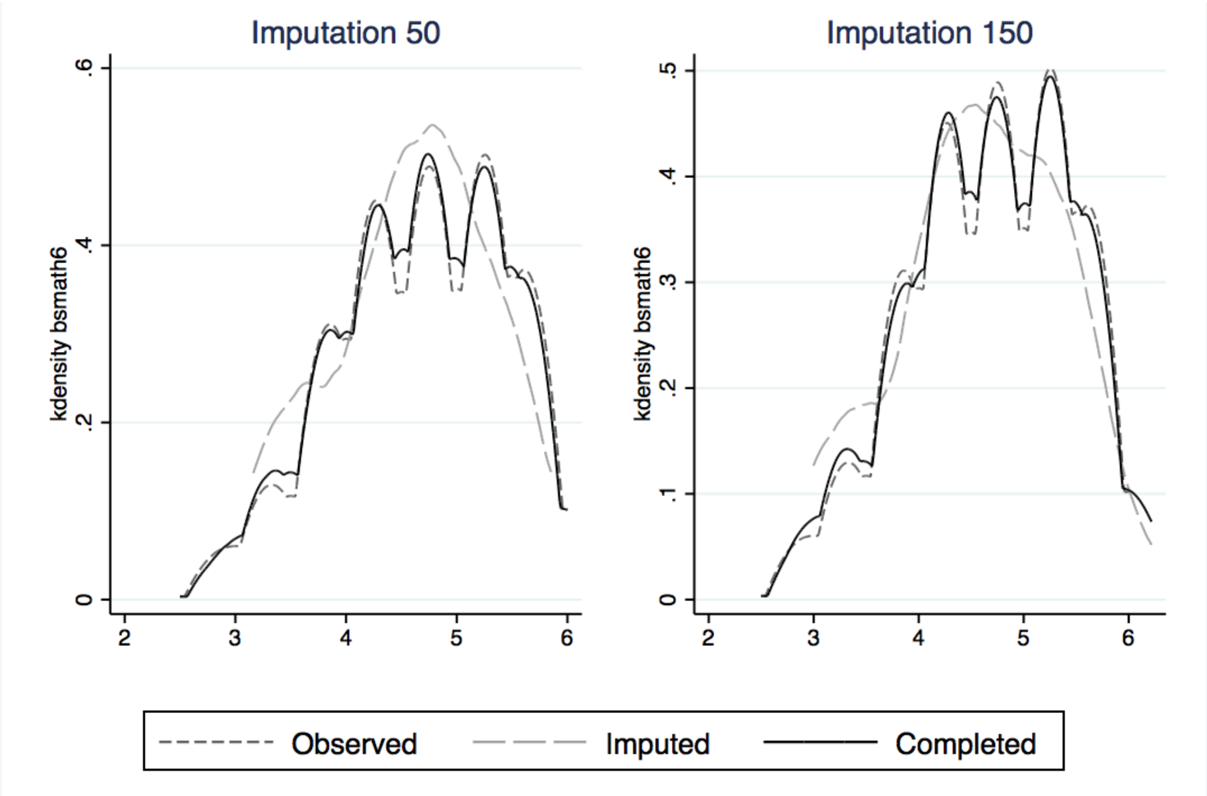
Figure A1: Traceplot of Iterations Math Grade 6th School Year



This holds true for all imputed variables. Nevertheless, we decided to use 50 iterations as robust inference is no longer obtained at the expense of increased estimation time. For the same reason, 200 imputed datasets were generated instead of the common number of less than 50 imputations.

The imputation process is quite satisfactory in the present case. All in all, the completed cases (i.e., the data after imputation) are quite similar to the complete cases regarding their central tendency, variation, and the quantiles (see Table A3 for the corresponding comparisons). Additionally, this conclusion is further supported by graphical comparisons of

Figure A2: Comparison of Imputed Values – Grade 6th School Year (Dependent Variable)



complete, imputed, and completed cases (e.g., Figure A2 for dataset 50 and 150; graphs for the other variables are available upon request). Where striking differences between the three occur (e.g., the higher share of people who are unemployed, in training or housekeepers – Figure A3 at the end of this section), they are in line with what we would expect from people's response behavior (e.g., in the mentioned case people with lowest class positions (unemployed) or who do not fit in one of the other higher ranked categories concerning occupational status (housekeepers, people in training) are less likely to respond – Kanuk & Berenson 1975: 448). Hence, overall the imputation process produces satisfactory results under the assumption that, conditional on observed characteristics, the missing pattern occurs randomly. At the same time, it should be stressed once again that parents' willingness to respond in the first place is far from being random but highly dependent on their social status (compare Table 6 in the main article; Proner 2011). As only the item-nonresponse for the sample of first wave respondents was imputed, the completed datasets provide additional statistical power and reliability for this selective sample. However, any inference beyond this selective sample should be avoided.

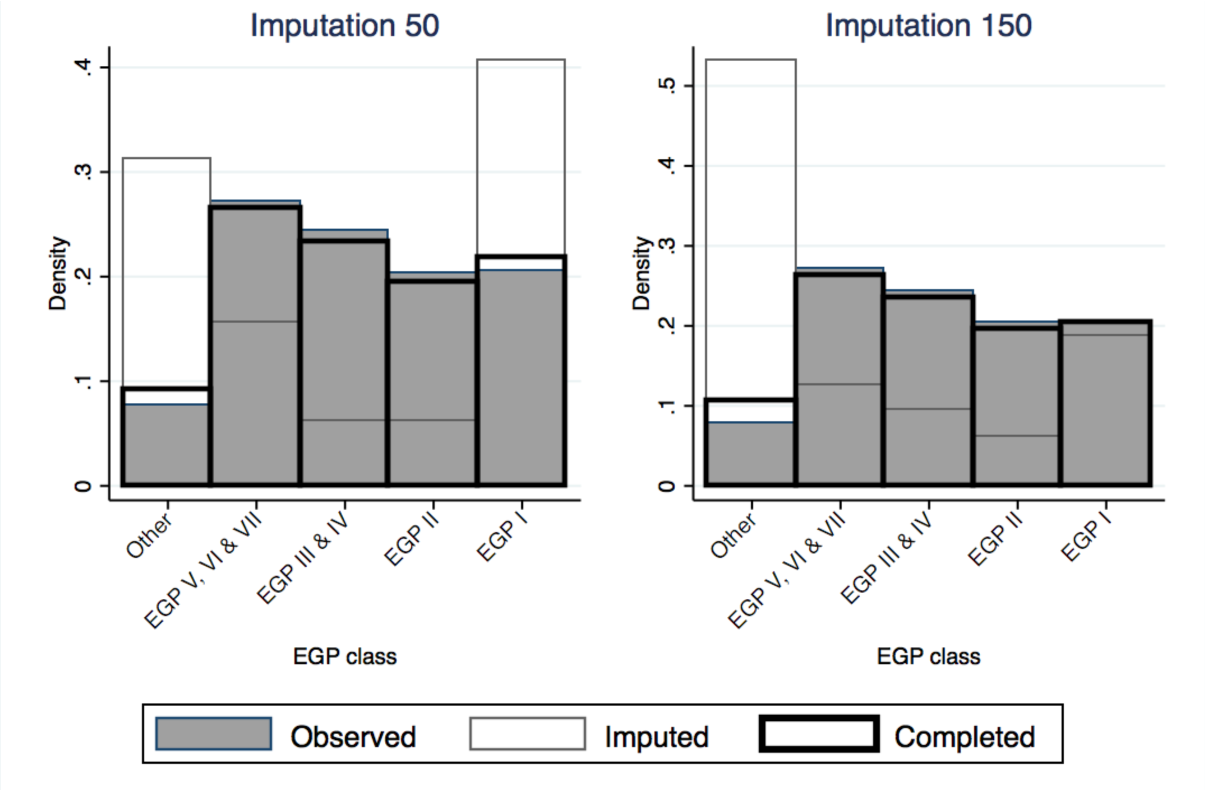
Table A3: Comparison of Descriptive Statistics

	Complete Cases (listwise deletion)				Imputed (M=200)	
	N	Mean (SD)	Min	Max	N	Mean (SD)
Dependent						
Grade (math) in 6th school year	353	4.756 (0.75)	3	6	525	4.676 (0.77)
Neighborhood characteristics						
Share (%) high status residents in neighborhood	353	10.897 (5.62)	3.5	23.4	525	10.386 (5.44)
Instrument						
Share (%) tenure with more than 5 rooms in neighborhood	353	9.374 (6.43)	1.21	27.64	525	9.056 (6.21)
Independent						
Continuous	N	Mean (SD)	Min	Max	N	Mean (SD)
Social integration	353	-0.008 (1.00)	-4.49	1.14	525	-0.002 (1.00)
Problem solving competency	353	0.011 (0.99)	-3.10	1.85	525	0.007 (1.01)
Perceived deviance	353	-0.018 (0.98)	-1.87	3.20	525	-0.002 (1.00)
Grade (math) 5th school year	353	4.712 (0.69)	2	6	525	4.624 (0.73)
Mean grade (math) class level	353	4.575 (0.33)	3.77	5.15	525	4.575 (0.33)

Categorical		listwise		Imputed	
	Values	N	Share	N	Share
Class position (EGP)	1 = Unemployed, in training, domestic work	353	6.23%	525	9.31%
	2 = EGP V, VI, VII		27.20%		26.81%
	3 = EGP III, IV		24.08%		23.42%
	4 = EGP II		19.55%		19.60%
	5 = EGP I		22.95%		20.86%
Language	0 = Other	353	13.31%	525	15.71%
	1 = German and other		45.89%		46.68%
	2 = German only		40.79%		37.61%
Sex	0 = Boys	353	46.18%	525	46.43%
	1 = Girls		53.82%		53.57%
Income situation	1 = Tense	353	15.30%	525	15.45%
	2 = Partly		30.31%		29.66%
	3 = Relaxed		54.39%		54.89%

Source: DEBIMISS, own calculations.

Figure A3: Comparison of Imputed Values - EGP Class Scheme



IV. Results from the Complete Case Analysis

In what follows, the results from the estimation models applied to the complete cases are outlined shortly. This enables further understanding of the impact and the consequences of the imputation process. As we shall see, the crucial relations in the data are reproduced and neither size nor significance of the effects are changed in an unexpected way. However, only a selection of all the models are presented and we will restrict the discussion on the theoretically important predictors (i.e., the neighborhood characteristics, the mediating variables and their interactions).

To begin with, Table A4 summarizes the main results of the OLS estimates for both, the epidemic and the collective socialization models. As the comparison with the coefficients from Table 2 and 3 in the main paper reveals, they all point in the same direction as the analyses based on the completed cases. While most relations differ only modestly in magnitude and significance (e.g., the estimates regarding the impact of the categorized share of high status residents in one's neighborhood according to the collective socialization hypothesis), a few exhibit quite striking differences between the complete and the completed cases. This is especially true for the interaction effect of the share of high status residents and the social integration into a local network for children with parents in the highest social classes (EGP I). Although the direction of the effects are maintained in the analysis with the imputed values, neither the interaction of the mediating variable with whether one lives in a neighborhood with 6 to 10%, 11 to 15% nor the one with the dummy for a neighborhood with more than 20% high status residents is significant. However, an (undocumented) analysis shows that neither the imputed values of the amount of social integration nor the ones of the dependent variable differ in their central tendency (mean) from the non-imputed ones. Hence, this pattern most likely emerges due to a complex interaction of other factors in the imputation process and must remain unanswered at this point. The crucial issue, however, is that the basic relationships in the OLS models are maintained.

The same holds true for the estimates of the IV models (compare Table A5 of this documentation with Table 4 in the paper). Although not all the significant effects of the neighborhood characteristics are reproduced using the 200 imputed datasets (e.g., the significant effect according to the epidemic theory for children with parents occupying skilled non-manual or self-employed class positions – EGP III and IV), the basic structure of association remains unaltered. Taken together, the OLS and the IV models suggest a small loss of statistical precision after a multiple imputation of the missing values. The imputation of crucial regressors, such as social origin, the mediating variables, or the prior mathematical achievement, seems to introduce further heterogeneity into the data. This explanation is supported by the generally increased standard errors of the estimates. Hence, although successful and valid, multiple imputation in the present case did not lead to increased statistical power. However, the imputation of missing values nevertheless seems justified as the increased heterogeneity also allows an alternative interpretation; ignoring the missing values would further restrict the interpretation of the results to an even more selective group of people.

Table A4: OLS Estimates Complete Case Analysis

	Model I: Epidemic Theory						Model II: Collective Socialization					
	Girls	Boys	EGP I	EGP II	EGP III, IV	EGP V-VII	Girls	Boys	EGP I	EGP II	EGP III, IV	EGP V-VII
Neighborhood characteristic												
5% or less high status residents	-0.047 (0.12)	-0.008 (0.07)	0.097 (0.26)	-0.057 (0.14)	-0.044 (0.08)	-0.057 (0.14)						
Share high status residents (Reference: max. 5%)												
6 to 10%							0.294* (0.14)	-0.096 (0.14)	0.371** (0.13)	0.234 (0.14)	-0.020 (0.15)	0.192 (0.20)
11 to 15%							0.129 (0.15)	-0.084 (0.15)	0.192 (0.16)	0.076 (0.10)	-0.128 (0.14)	0.145 (0.20)
16 to 20%							0.080 (0.41)	-0.161 (0.18)	1.120* (0.52)	0.112 (0.22)	-0.143 (0.29)	-0.732*** (0.15)
More than 20%							-0.018 (0.21)	-0.234 (0.19)	-0.011 (0.23)	0.079 (0.17)	-0.490 (0.31)	-0.007 (0.30)
Perceived deviance	0.020 (0.05)	0.001 (0.06)	-0.058 (0.04)	0.046 (0.10)	0.064 (0.04)	0.046 (0.10)						
Social integration							-0.107 (0.07)	-0.035 (0.14)	-1.681*** (0.40)	-0.163*** (0.03)	0.071 (0.21)	-0.076 (0.19)
Interactions												
Below threshold * Perceived deviance	-0.016 (0.09)	0.063 (0.08)	0.371 (0.35)	-0.077 (0.12)	-0.047 (0.06)	-0.077 (0.12)						
Social integration * Share high status												
6 to 10%							0.131 (0.08)	-0.023 (0.14)	1.705*** (0.41)	0.259** (0.08)	-0.038 (0.22)	0.065 (0.20)
11 to 15%							0.112 (0.07)	0.183 (0.16)	1.835*** (0.39)	0.208*** (0.05)	-0.148 (0.23)	0.161 (0.21)
16 to 20%							0.156 (0.56)	0.331* (0.14)	-0.085 (0.86)	0.495*** (0.05)	0.355 (0.40)	0.000 (.)
More than 20%							0.003 (0.12)	0.023 (0.15)	1.703*** (0.41)	0.041 (0.09)	-0.065 (0.20)	-0.452 (0.69)
Sex (Reference: Boy)			0.084 (0.08)	0.124 (0.11)	-0.085 (0.10)	0.124 (0.11)			0.047 (0.07)	0.067 (0.09)	-0.082 (0.13)	0.174^{+} (0.09)
Grade 4th school year (Math)	0.661*** (0.07)	0.816*** (0.07)	0.816*** (0.06)	0.718*** (0.09)	0.726*** (0.05)	0.718*** (0.09)	0.679*** (0.06)	0.804*** (0.08)	0.804*** (0.07)	0.710*** (0.09)	0.721*** (0.07)	0.740*** (0.09)
Mean grade class level (Math)	0.444** (0.12)	0.242* (0.10)	0.466* (0.17)	0.353* (0.13)	0.380^{+} (0.18)	0.353* (0.13)	0.619** (0.18)	0.305* (0.14)	0.891** (0.26)	0.137 (0.35)	0.620^{+} (0.32)	0.528** (0.15)
R^2	0.634	0.704	0.657	0.563	0.683	0.563	0.657	0.729	0.696	0.632	0.706	0.595
N	190	163	81	96	85	96	190	163	81	69	85	96

Dependent variable: Math grade in 6th school year; additionally controlled for social origin (class position), language prior to enrollment, problem solving skills, and financial situation; standard errors (corrected for heteroscedasticity at neighborhood level) in parentheses. Source: DEBIMISS, own calculations. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: IV Estimates Complete Case Analysis

	Model I: Epidemic Theory						Model II: Collective Socialization					
	Girls	Boys	EGP I	EGP II	EGP III, IV	EGP V-VII	Girls	Boys	EGP I	EGP II	EGP III, IV	EGP V-VII
Neighborhood characteristic												
5% or less high status residents	0.285	0.110	-0.033	0.232	0.372**	0.502*						
	(0.18)	(0.23)	(0.16)	(0.24)	(0.13)	(0.20)						
Share high status residents							-0.021	-0.026^{+}	-0.022	-0.005	-0.044	-0.082*
							(0.01)	(0.02)	(0.01)	(0.02)	(0.04)	(0.04)
Perceived deviance	0.020	0.016	-0.041	-0.028	0.048	0.018						
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.08)						
Social integration							-0.004	0.056	0.102*	0.027	0.015	0.023
							(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.06)
Sex (Reference: Boy)			0.079	0.006	-0.091	0.093			0.070	0.050	-0.154	0.116
			(0.08)	(0.07)	(0.09)	(0.10)			(0.06)	(0.06)	(0.11)	(0.10)
Grade 4th school year (Math)	0.660***	0.808***	0.808***	0.657***	0.739***	0.732***	0.664***	0.782***	0.774***	0.680***	0.732***	0.733***
	(0.06)	(0.07)	(0.06)	(0.10)	(0.05)	(0.09)	(0.06)	(0.08)	(0.05)	(0.10)	(0.06)	(0.09)
Mean grade class level (Math)	0.588***	0.312^{+}	0.519***	0.106	0.589**	0.676***	0.679***	0.591**	0.898**	0.087	0.847^{+}	0.836***
	(0.14)	(0.16)	(0.15)	(0.23)	(0.19)	(0.16)	(0.20)	(0.20)	(0.28)	(0.42)	(0.48)	(0.25)
Rho / F	-0.558*	-0.206	0.026	-0.228	-0.829***	-0.884**	0.345	2.366	0.100	0.207	1.123	4.826*
p	0.013	0.558	0.876	0.627	0.000	0.002	0.564	0.140	0.756	0.655	0.303	0.041
N	190	163	81	69	85	96	190	163	81	69	85	96

Dependent variable: Math grade in 6th school year; instrument: share tenure with more than 5 rooms in neighborhood; additionally controlled for social origin (class position), language prior to enrollment, problem solving skills, and financial situation; standard errors (corrected for heteroscedasticity at neighborhood level) in parentheses.
Source: DEBIMISS, own calculations. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

V. Additional tables

Table A6: Descriptive Statistics (Imputed Data)

	N	Mean (SD)	Min	Max	% imputed
Dependent					
Grade (math) in 6th school year	525	4.676 (0.77)	3	6	17.14%
Neighborhood characteristic					
Share (%) high status residents in neighborhood	525	10.386 (5.44)	3.5	23.4	0%
Instrument					
Share (%) tenure with more than 5 rooms in neighborhood	525	9.056 (6.21)	1.21	27.64	0%
Independent					
Continuous	N	Mean (SD)	Min	Max	
Social integration	525	-0.002 (1.00)	-4.49	1.77	2.86%
Problem solving competencies	525	0.007 (1.01)	-3.15	2.79	16.38%
Perceived deviance	525	-0.002 (1.00)	-2.37	3.85	7.24%
Grade (Math) 5th school year	525	4.624 (0.73)	2	6	2.67%
Mean grade (Math) class level	525	4.575 (0.33)	3.77	5.23	2.48%

Categorical	Values	N	Share	% Imputed
Class position (EGP)	1 = Unemployed, in training, domestic work 2 = EGP V, VI, VII 3 = EGP III, IV 4 = EGP II 5 = EGP I	525	9.31% 26.81% 23.42% 19.60% 20.86%	6.10%
Language	0 = Other 1 = German and other 2 = German only	525	15.71% 46.68% 37.61%	1.14%
Sex	0 = Boys 1 = Girls	525	46.43% 53.57%	1.14%
Income situation	1 = Tense 2 = Partly 3 = Relaxed	525	15.45% 29.66% 54.89%	1.14%

Table A7: Social Composition of Neighborhoods (complete cases)

Share high status in neighborhood	Other	EGP V-VII	EGP III, IV	EGP II	EGP I	Total (N)
5% or less	6.25%	31.25%	37.5%	12.5%	12.5%	100% (16)
6 - 10%	7.59%	37.24%	26.21%	17.93%	11.03%	100% (145)
11 - 15%	5.93%	24.44%	22.96%	16.3%	30.37%	100% (135)
16 - 20%	0.00%	9.09%	27.27%	45.45%	18.18%	100% (11)
More than 20%	4.35%	6.52%	15.22%	30.43%	43.48%	100% (46)
Total	6.23%	27.2%	24.08%	19.55%	22.95%	100% (353)

Table A8: Construction of Relevant Scales

Items & Scales	N	Mean	SD	Factor Score	Cronbachs' α (if item is deleted)
Tease teacher	515	2.371	1.08	0.753	0.710
Destroy things	515	1.816	0.87	0.643	0.733
Beat other's up	519	2.860	1.11	0.580	0.750
Tease schoolfellows	517	2.221	1.00	0.777	0.700
Being rude to teacher	516	3.132	1.13	0.704	0.719
Disrupt classroom	509	1.527	0.78	0.502	0.757
Skip school	518	1.411	0.65	0.496	0.759
Perceived deviance[a]	**487**	**0**	**1**	**0.807**[c]	**0.762**
I feel accepted	522	4.301	0.97	0.792	0.641
I feel excluded at school	522	4.435	1.05	0.687	0.694
It is easy to make friends	522	3.992	1.18	0.657	0.700
I feel lonely at school	522	4.686	0.78	0.744	0.681
I am popular at my school	517	2.965	1.24	0.639	0.710
Social integration[b]	**510**	**0**	**1**	**0.775**[c]	**0.731**
Master problems with own strength	441	3.644	0.87	0.793	0.742
Whatever happens, I'll be alright	439	3.613	0.96	0.843	0.704
I can rely on myself	441	3.562	1.03	0.813	0.726
I always find a solution	441	3.624	0.96	0.695	0.793
Problem solving competencies[b]	**439**	**0**	**1**	**0.784**[c]	**0.794**

[a] [1 "Happens never" … 5 "Happens all the time"]
[b] [1 "Completely disagree" … 5 "Completely agree"]
[c] KMO of one-factor solution

References

Allison, P.D., 2001: Missing Data. Quantitatives Apllications in the Social Sciences 136. Thousand Oaks: Sage.

Carpenter, J.R. & M.G. Kenward, 2013: Multiple Imputation and Its Application. Statistics in Practise. New York: Wiley.

Kanuk, L. & C. Berenson, 1975: Mail Surveys and Response Rates: A Literature Review. Journal of Marketing Research 12: 440–453.

McKnight, P.E., K.M. McKnight, S. Sidani & A.J. Figueredo, 2007: Missing Data: A Gentle Introduction. New York & London: Guilford.

Proner, H., 2011: Ist Keine Antwort Auch Eine Antwort? Wiesbaden: VS.

Rubin, D.B., 1987: Multiple Imputation for Nonresponse in Surveys. Hoboken: Wiley.

StataCorp, 2013: Stata Release 13. Vol. 20: Multiple Imputation. College Station: Stata Press.

Von Hippel, P.T., 2009: How to Impute Interactions, Squares, and Other Transformed Variables. Sociological Methodology 39: 265–291.

White, I.R., P. Royston & A.M. Wood, 2011: Multiple Imputation Using Chained Equations: Issues and Guidance for Practice. Statistics in Medicine 30: 377–399.