Do pupils at research-informed schools actually perform better? Findings from a study at English schools

Jana Groß Ophoff, Chris Brown, Christoph Helm

Appendix: Technical details

To begin with, the evaluation of the measurement models for each scale is reported for level 1 (individual teachers) and level 2 (school). This is followed by the results for the structural model, which was derived from the theoretically grounded assumptions in previous chapters (see Figure 1). In conclusion, the average self-reported research use etc. is compared between teachers and schools with different Ofsted-judgements.

A1: Evaluation of measurement models

Single-level and multilevel confirmatory factor analyses demonstrate acceptable to good fit for all models (Table 2), which indicates that the proposed measurement models reasonably account for the observed data. The only exception is the level 2-SRMR for OL, which is too high.

Table A1: Confirmatory factor analyses of the study variables

Model	# items	# par	Chi2	df	Chi2/df	CFI	TLI	RMSEA	SRMR
Single-level analysis									
Trust	6		28.948	8	3.6	.959	.924	.057	.033
Organizational learning	6	18	34.890	9	3.9	.978	.964	.060	.025
Research use	7	22	73.783	13	5.7	.929	.885	.073	.045
Multilevel analysis (shared constructs)									
Trust	6	38	20.096	10	2.1	.988	.963	.036	.005/.096
Organizational learning	6	38	6.268	10	0.6	1.000	1.006	.000	.002/.215
Research use	7	49	14.938	14	1.1	.999	.997	.009	.002/.061

Note. Abbreviations: # items = number of included items; # par = number of model parameters; df = degrees of freedom.

A2: Comparison of the factor structure for the Research Use scale and the Organizational Learning scale

The high level 2 correlations between the Research Use (RU) scale and the Organizational Learning (OL) scale raise the question whether it is warranted at all to empirically distinguish between the two constructs on school level. Table 4 sheds light on the question, whether RU and OL should be modeled as separate constructs or not. The comparison of the factor structure at school level shows that model fit significantly increases when RU and OL are modelled as joint single factor (i.e., the correlation between the two factors is constrained to 1) instead of two separate factors. However, the same does not apply on single-level, where the model with two separate factors proves to be better fitting to the data.

Table A2: Comparison of the factor structure for the Research Use scale and the Organizational Learning scale on single-level and multilevel analysis

Levels of Analysis	# factors	# par	Chi2	df	Chi2/df	CFI	TLI	RMSEA	SRMR	corre-lation	p	Δ Chi2	p
Single-level analysis	2: RU + OL	40	290.939	64	4.5	.912	.893	.063	.049	.169	.001		
	1: RUOL	39	1054.213	65	16.2	.617	.540	.131	.136			-763.274	.001
Multilevel analysis	2: RU + OL	131	92.731	64	1.4	.992	.981	.022	.003/.290	.953	.001		
	1: RUOL	130	88.593	65	1.4	.994	.984	.020	.003/.289			4.148	.042

Note. Abbreviations: # factors = number of factors; # par = number of model parameters; df = degrees of freedom.

A3: Measurement invariance testing

As the effect of different school inspection ratings is of interest, too, the measurement invariance was analyzed for each of the three self-assessment scales. Due to the small number of "requires improvement" schools (N = 3), only the comparability of "good" vs. "outstanding" schools was examined, for which the results are reported in Table 5. However, only partial invariance could be established, which means that some coefficients had to be estimated freely across groups in order to obtain a non-significant drop in model fit. For some items of the TR scale (items 2, 3, 4, see Table 1) and the RU scale (items 2, 4, 5), only configural invariance is to be assumed, as the factor loadings had to be estimated freely across groups in order to obtain a non-significant drop in model fit. Furthermore, for the OL scale the intercepts differ between groups for the items 2, 3, and 4 (see Table 1).

Table A3: Measurement invariance testing (good vs. outstanding schools on level-1)

Model	Measurement invariance	df	AIC	BIC	Chi2	Δ Chi2	Δ df	p
TR	configural	18	7634.726	7798.513	185.246			
TR	weak (metric)	20	7633.849	7788.537	188.369	3.123	2	.210
TR	strong (scalar)	25	7631.589	7763.528	196.109	7.74	5	.171
OL	configural	18	10832.258	10995.994	71.484			
OL	weak (metric)	23	10824.976	10965.971	74.202	2.718	5	.743
OL	strong (scalar)	25	10825.134	10957.032	78.360	4.158	2	.125
RU	configural	28	12085.532	12278.726	223.846			
RU	weak (metric)	32	12082.178	12256.973	228.492	4.646	4	.326
RU	strong (scalar)	38	12079.633	12226.829	237.947	9.455	6	.150

Note. Note. Abbreviations: TR = trust scale (see 4.2.1); OL = Organizational learning scale (see 4.2.3); RU = Research use climate scale (see 4.2.4). Only partial invariance could be established: Factor loadings of TR2, TR4, TR5, RU4, RU5, and intercepts of OL2, OL3, OL4 (item wording, see Table 1).