ZENTRUM FÜR INTERNATIONALE BILDUNGSVERGLEICHSSTUDIEN

IPN Leibniz Institute for Science and Mathematics Education

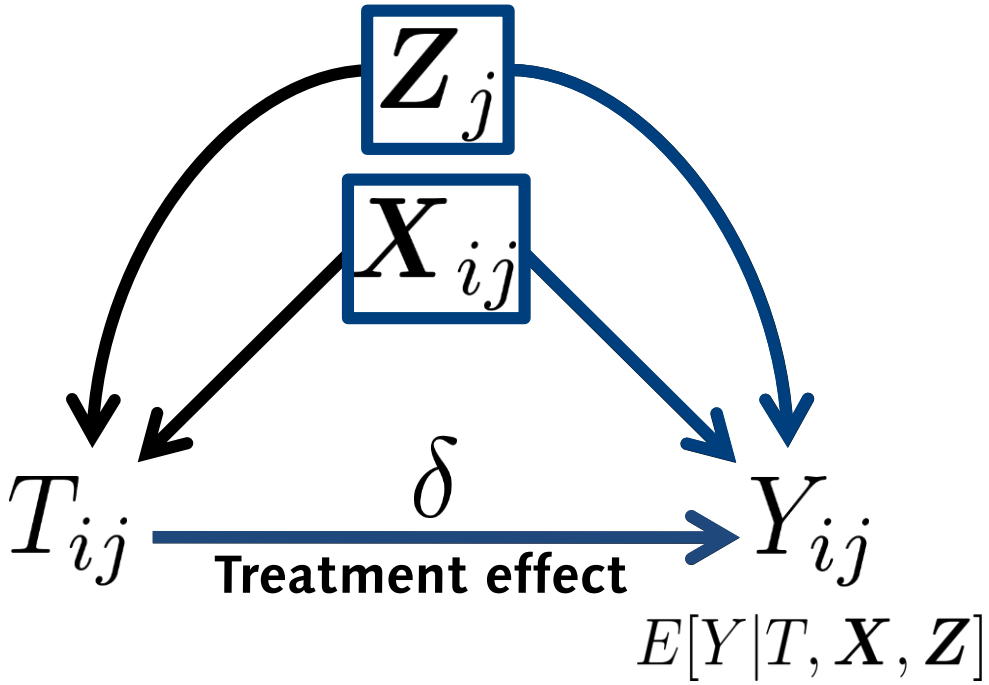
Multilevel Propensity Scores: An Evaluation of Findings

Alvaro Fuentes[1] (fuentes@ipn.uni-kiel.de), Oliver Lüdtke[12], Alexander Robitzsch[12]

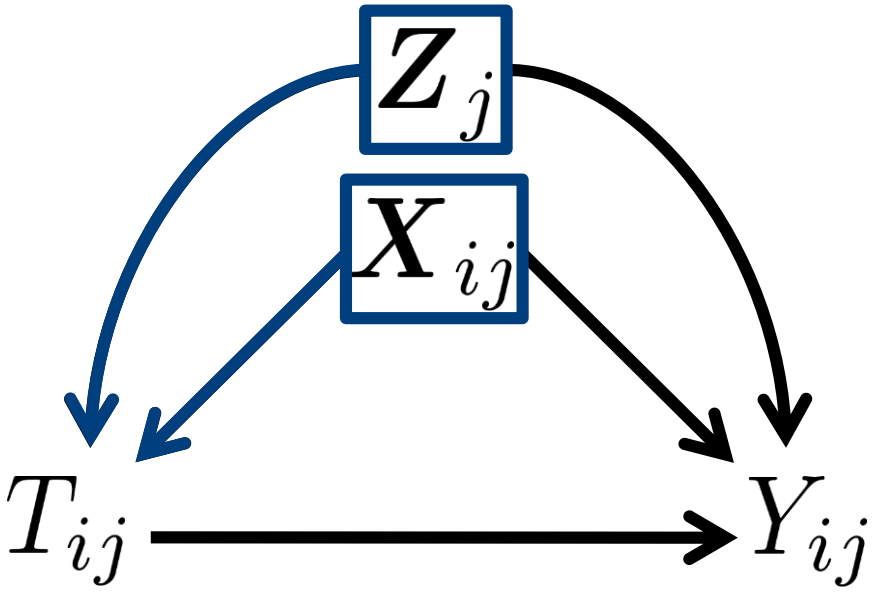
[1] Leibniz Institute for Science and Mathematics Education

[2] Center for International Student Assessment

Treatment effect estimation



Modelling treatment assignment



$E[T|\boldsymbol{X},\boldsymbol{Z}]$

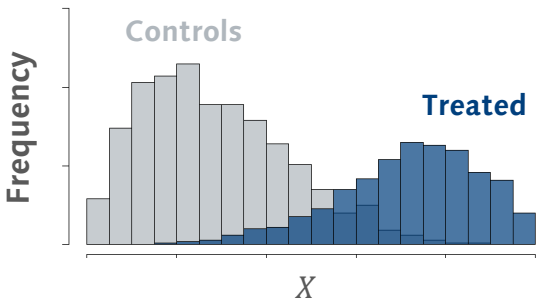
$Pr(T = 1|\boldsymbol{X},\boldsymbol{Z})$

Propensity score

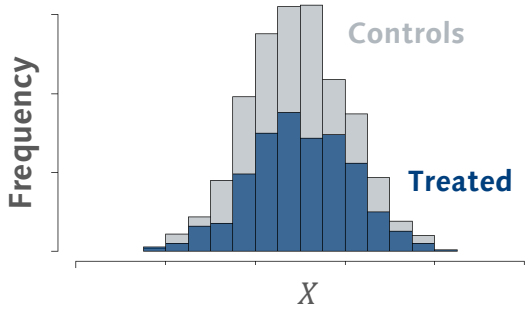
$e(\boldsymbol{X},\boldsymbol{Z})$

Propensity score weighting

Observational study



Randomized experiment



Balancing weights (Li et al., 2018)

Distribution among the treated

$$p_1(X) \propto p(X)e(X)$$

Distribution among the controls

$$p_0(X) \propto p(X)(1-e(X))$$

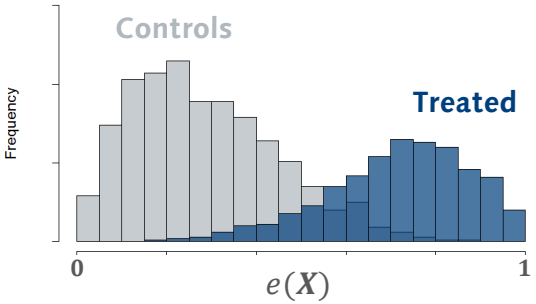
Name	Weights
Inverse Probability of Treatment Weights (IPTW)	$\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$
Overlap Weights	$(1-e(X), e(X))$
Truncated Weights	$\left(\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{e(X)}, \frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{1-e(X)}\right)$

Balancing weights (Li et al., 2018)

$$\frac{\mathbf{1}}{\boldsymbol{e(X)}} p_1(X) \propto p(X)\cancel{e(X)}$$

$$\frac{\mathbf{1}}{\mathbf{1-}\boldsymbol{e(X)}} p_0(X) \propto p(X)\cancel{(1-e(X))}$$

Name	Weights
Inverse Probability of Treatment Weights (IPTW)	$\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$
Overlap Weights	$(1-e(X), e(X))$
Truncated Weights	$\left(\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{e(X)}, \frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{1-e(X)}\right)$

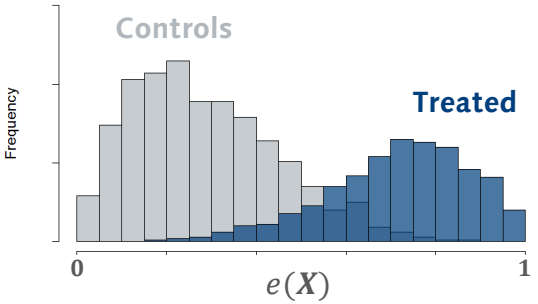


Balancing weights (Li et al., 2018)

$$(\mathbf{1} - \boldsymbol{e}(\boldsymbol{X}))\, p_1(X) \propto p(X)e(X)\,(\mathbf{1} - \boldsymbol{e}(\boldsymbol{X}))$$

$$\boldsymbol{e}(\boldsymbol{X})\, p_0(X) \propto p(X)(1 - e(X))\,\boldsymbol{e}(\boldsymbol{X})$$

Name	Weights
Inverse Probability of Treatment Weights (IPTW)	$\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$
Overlap Weights	$(1 - e(X), e(X))$
Truncated Weights	$\left(\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{e(X)}, \frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{1-e(X)}\right)$

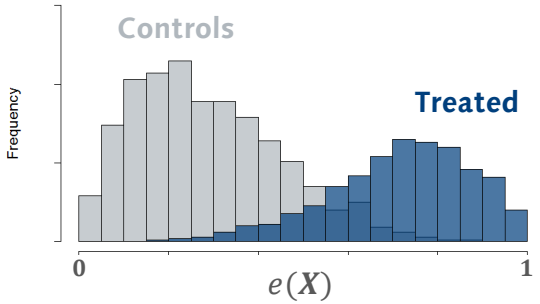


Balancing weights (Li et al., 2018)

$$\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{e(X)}\, p_1(X) \propto p(X)e(X)\,\mathbf{1}(\alpha < e(X) < 1-\alpha)$$

$$\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{1-e(X)}\, p_0(X) \propto p(X)(1-e(X))\,\mathbf{1}(\alpha < e(X) < 1-\alpha)$$

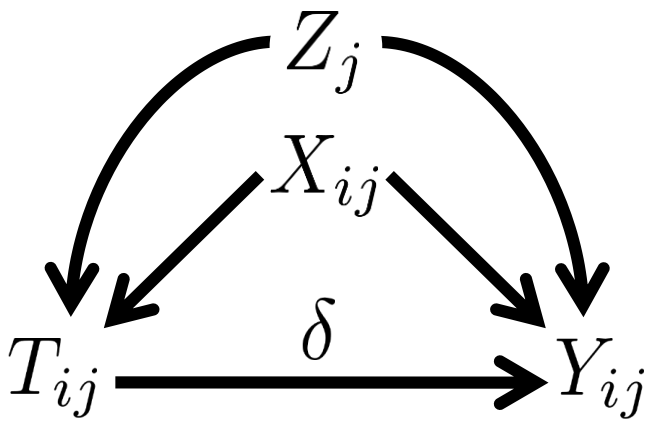
Name	Weights
Inverse Probability of Treatment Weights (IPTW)	$\left(\frac{1}{e(X)}, \frac{1}{1-e(X)}\right)$
Overlap Weights	$(1-e(X), e(X))$
Truncated Weights	$\left(\frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{e(X)}, \frac{\mathbf{1}(\alpha < e(X) < 1-\alpha)}{1-e(X)}\right)$



Multilevel PS literature findings

- Asymptotically unbiased estimation conditioning both across and within clusters (Li, Zaslavsky & Landrum, 2013; for PS matching, Steiner, Kim & Thoemmes, 2012)
- Underperformance of RE propensities vs FE propensities (Thoemmes & West, 2011; Li, Zaslavsky & Landrum, 2013)
- Automatic conditioning on omitted context (Arpino & Mealli, 2011; Li, Zaslavsky & Landrum, 2013)

Simulation study design



$$\text{logit}(e_{ij}) = \alpha_X X_{ij} + \alpha_Z Z_j + u_{Tj}$$
$$Y_{ij} = \delta T_{ij} + \beta_X X_{ij} + \beta_Z Z_j + u_{Yj} + \epsilon_{Yij}$$

Number of clusters (N)	30
Units per cluster (n_j)	10, 12,...,100
Outcome equation R^2	.3 ($\delta = \beta_X = \beta_Z = 0.49$)
Treatment equation R^2	.2 or .6
Residual ICC of T_{ij}	.1 or .4
Residual ICC of Y_{ij}	.2
Proportion treated	.5

Normally distributed residuals u_{Tj}, u_{Yj}, ϵ_{Yij}

Propensity score estimators

$$\text{logit}(\hat{e}_{ij}) = \hat{\alpha}_{0j} + \hat{\alpha}_1 x_{ij}$$

- **Fixed effects (FE) model**
  Cluster-specific intercepts $\hat{\alpha}_{0j}$ based on dummy indicators
  → No distributional assumption
- **Random effects (RE) model**
  Multilevel logistic regression model
  Cluster-specific intercepts $\hat{\alpha}_{0j}$ assumed to follow a normal distribution
  Propensity score based on empirical Bayes estimate

Treatment effect estimators

- **Across clusters**
  Weighted mean difference with the entire sample

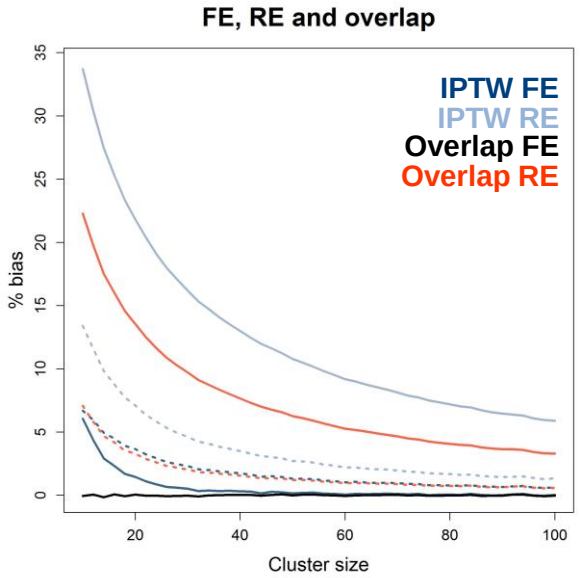
$$\hat{\delta}_{across} = \frac{\sum_{j=1}^{N}\sum_{i=1}^{n_j} t_{ij} w_{ij} y_{ij}}{\sum_{j=1}^{N}\sum_{i=1}^{n_j} t_{ij} w_{ij}} - \frac{\sum_{j=1}^{N}\sum_{i=1}^{n_j} (1-t_{ij}) w_{ij} y_{ij}}{\sum_{j=1}^{N}\sum_{i=1}^{n_j} (1-t_{ij}) w_{ij}}$$

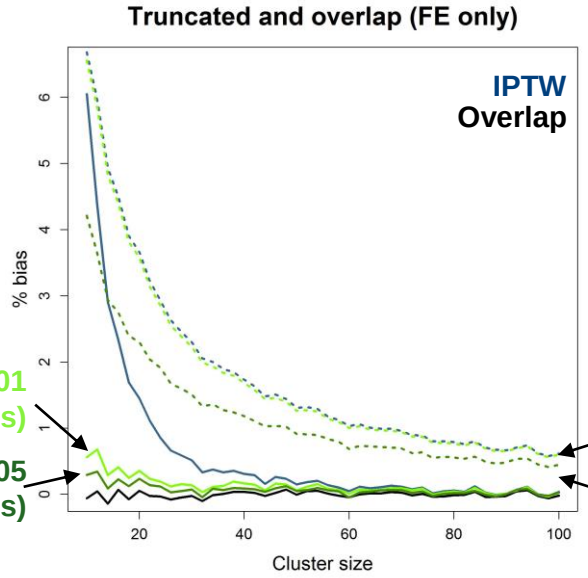
- **Within clusters**
  Weighted average of cluster-specific weighted mean differences

$$\hat{\delta}_{j} = \frac{\sum_{i=1}^{n_j} t_{ij} w_{ij} y_{ij}}{\sum_{i=1}^{n_j} t_{ij} w_{ij}} - \frac{\sum_{i=1}^{n_j} (1-t_{ij}) w_{ij} y_{ij}}{\sum_{i=1}^{n_j} (1-t_{ij}) w_{ij}}$$

$$\hat{\delta}_{within} = \frac{\sum_{j=1}^{N} w_{\bullet j} \hat{\delta}_j}{\sum_{j=1}^{N} w_{\bullet j}} \quad \text{, where} \quad w_{\bullet j} = \sum_{i=1}^{n_j} w_{ij}$$

Results





Truncated 0.01 (across)

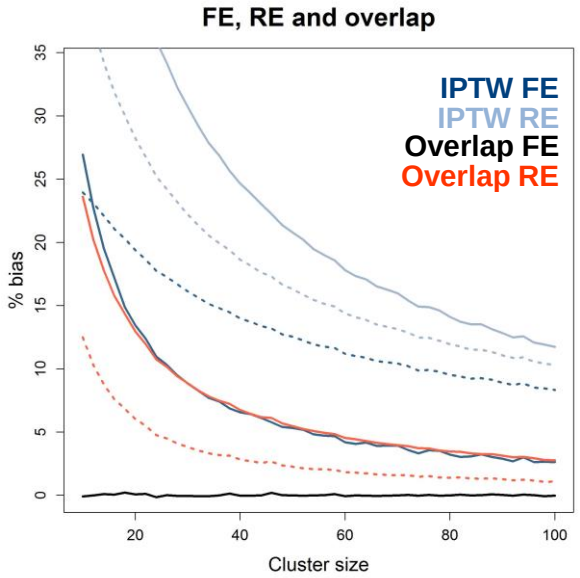
Truncated 0.05 (across)

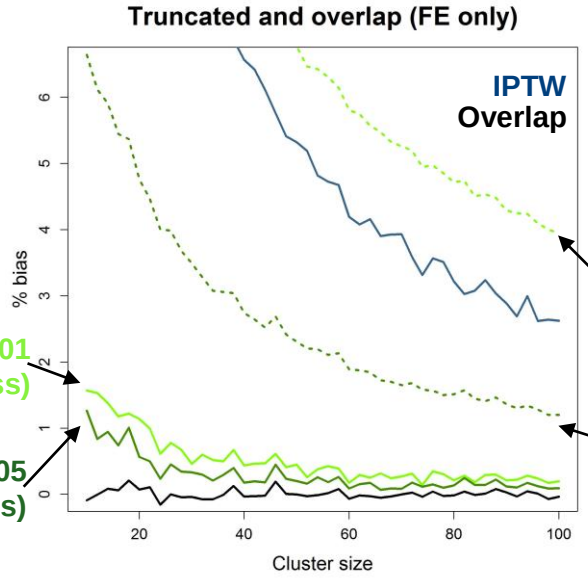
Truncated 0.01 (within)

Truncated 0.05 (within)

Weights		$n_j = 10$ % Bias	$n_j = 10$ RMSE	$n_j = 50$ % Bias	$n_j = 50$ RMSE
IPTW FE	across	6.05%	0.214	0.15%	0.072
	within	6.68%	0.205	1.31%	0.071
Truncated FE α=0.05	across	0.29%	0.157	0.05%	0.066
	within	4.21%	0.170	0.91%	0.066
Overlap FE	across	-0.06%	0.138	0.00%	0.059
	within	-0.06%	0.138	0.00%	0.059
IPTW RE	across	33.68%	0.706	10.78%	0.231
	within	13.39%	0.310	2.72%	0.085
Truncated RE α=0.05	across	32.48%	0.681	9.93%	0.213
	within	13.00%	0.302	2.43%	0.080
Overlap RE	across	22.26%	0.477	6.26%	0.141
	within	7.07%	0.200	1.21%	0.064

Treatment equation $R^2 = .6$





Truncated 0.01 (within)

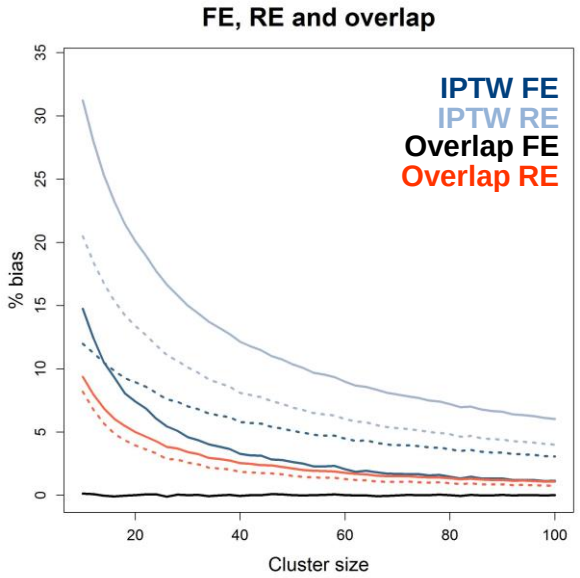
Truncated 0.01 (across)

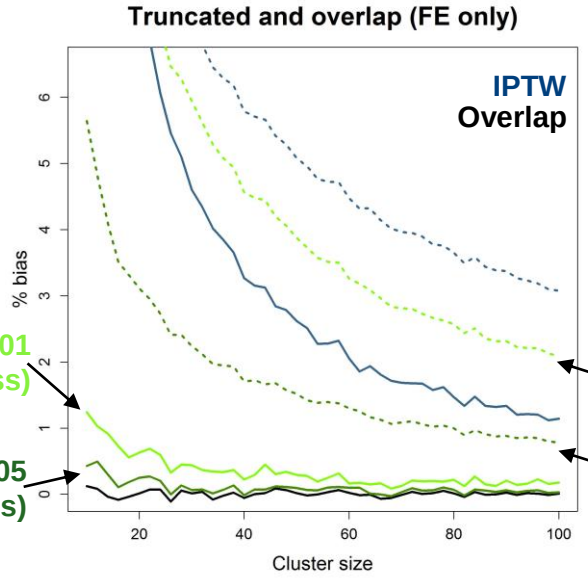
Truncated 0.05 (within)

Truncated 0.05 (across)

Weights		$n_j = 10$ % Bias	$n_j = 10$ RMSE	$n_j = 50$ % Bias	$n_j = 50$ RMSE
IPTW FE	across	26.94%	0.656	5.32%	0.234
	within	23.96%	0.553	12.55%	0.287
Truncated FE α=0.05	across	1.26%	0.211	0.20%	0.088
	within	6.64%	0.232	2.30%	0.095
Overlap FE	across	-0.09%	0.171	0.00%	0.072
	within	-0.09%	0.171	0.00%	0.072
IPTW RE	across	64.46%	1.342	20.78%	0.458
	within	40.49%	0.853	16.31%	0.353
Truncated RE α=0.05	across	32.34%	0.687	7.11%	0.169
	within	22.49%	0.493	5.16%	0.134
Overlap RE	across	23.60%	0.509	5.46%	0.132
	within	12.49%	0.305	2.26%	0.085

Treatment residual ICC = .4





Truncated 0.01 (across)

Truncated 0.05 (across)

Truncated 0.01 (within)

Truncated 0.05 (within)

Weights		n_j = 10 % Bias	n_j = 10 RMSE	n_j = 50 % Bias	n_j = 50 RMSE
IPTW FE	across	14.73%	0.373	2.62%	0.128
	within	11.99%	0.307	5.08%	0.140
Truncated FE α=0.05	across	0.43%	0.185	0.09%	0.079
	within	5.64%	0.204	1.52%	0.081
Overlap FE	across	0.12%	0.157	0.02%	0.066
	within	0.12%	0.157	0.02%	0.066
IPTW RE	across	31.21%	0.663	10.36%	0.233
	within	20.49%	0.453	6.93%	0.166
Truncated RE α=0.05	across	18.02%	0.404	3.31%	0.102
	within	17.06%	0.387	3.40%	0.102
Overlap RE	across	9.36%	0.242	2.12%	0.079
	within	8.18%	0.229	1.54%	0.073

Summary

- Conditioning across clusters may be sufficient if the propensity score is correctly specified.
- FE propensities dominate RE in all the simulated conditions in terms of bias and RMSE.
- Using propensities with cluster-specific intercepts we can automatically control for any omitted context. When clusters are small, FE propensities are best suited for this.
- Truncated and overlap weights are a practical and effective correction for the bias of extreme propensities also in the multilevel data case.

Next steps: more complex data-generating processes (interactions, random slopes…)

ZENTRUM FÜR INTERNATIONALE BILDUNGSVERGLEICHSSTUDIEN

IPN Leibniz Institute for Science and Mathematics Education

Thank you

Alvaro Fuentes[1] (fuentes@ipn.uni-kiel.de), Oliver Lüdtke[12], Alexander Robitzsch[12]
[1] Leibniz Institute for Science and Mathematics Education
[2] Center for International Student Assessment