UiO : CEMO – Centre for Educational Measurement
University of Oslo

Synthesizing research on complex sampling surveys: Two-stage meta-analysis with individual participant data

Diego G. Campos, Mike W.-L. Cheung, & Ronny Scherer

Centre for Educational Measurement at the University of Oslo
and National University of Singapore

IMPS, Bologna
12.07.2022

Individual Participant Data (IPD) Meta-Analysis

Pooling raw data

Comparability of analytic models

Sample and measurement heterogeneity

Decreased risk of ecological bias

Average effects

Quantify heterogeneity

Moderator effects

Complex Survey Data

1. Multi-stage sampling
2. Survey weights
3. Rotated Questionnaire Designs

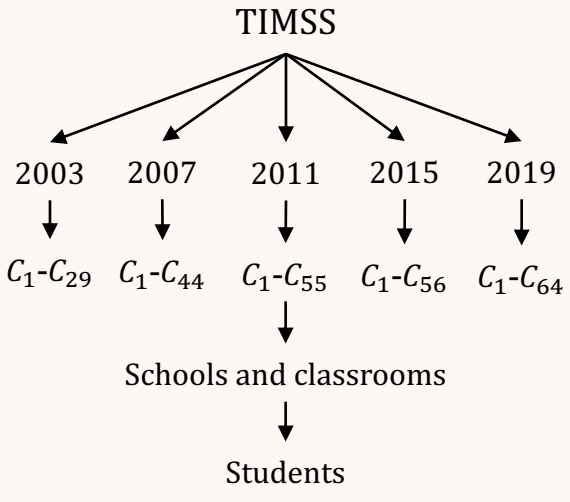
Examples

- Program for International Student Assessment (PISA)
- Trends in International Mathematics and Science Study (TIMSS)
- National Assessment of Educational Progress (NAEP)
- National Educational Panel Study (NEPS)

Rutkowski et al. (2010)

Two-Stage IPD Meta-Analysis

Stage 1: Raw Data Analysis



- Hierarchical structure
- Sampling weights
- Plausible values
- Measurement Invariance

i: Students, *j*: Classrooms, *k*: Countries

Stage 2: Meta-Analysis

- Multivariate meta-analysis
- Multilevel meta-analysis
- Mixed-effects meta-regression

1. Problem formulation and framing of the meta-analysis

Generalizability of the Big-Fish-Little-Pond-Effect (BFLPE) across countries and TIMSS waves

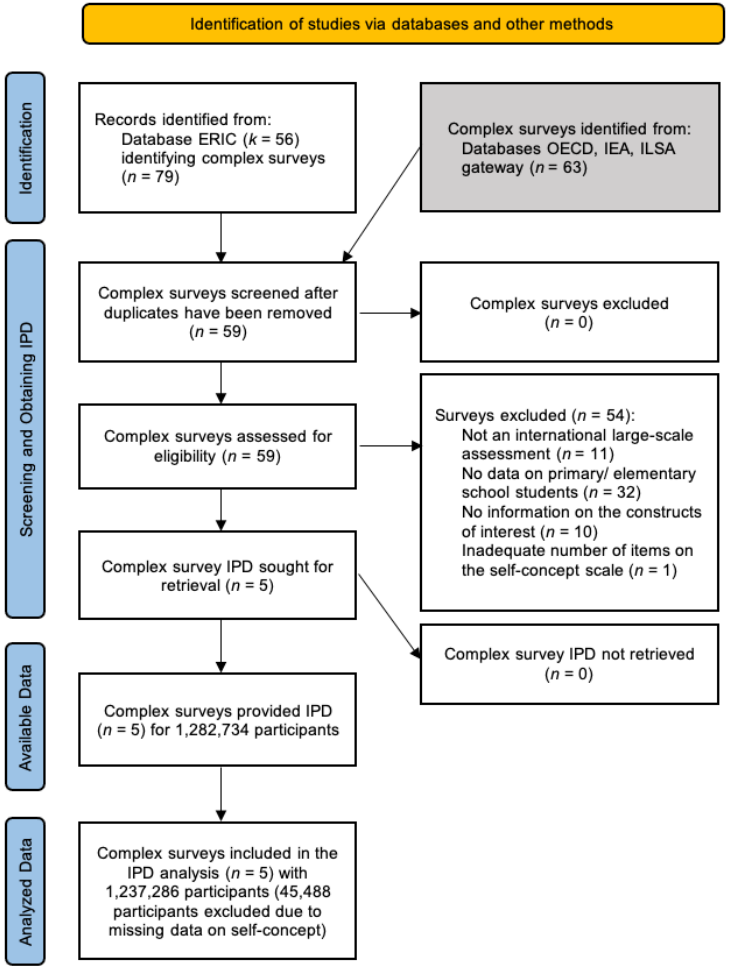
RQ1: Evidence supporting the generalizability of the BFLPE across countries and studies?

RQ2: Explanation of heterogeneity by country-level variables (e.g., Human development index, power distance?

2. Data search and selection

Search terms:
"International large-scale assessment" AND "Mathematics achievement" AND "Self-concept"

Inclusion/Exclusion Criteria:
- Mathematics self-concept
- Mathematics achievement
- Classroom level comparisons

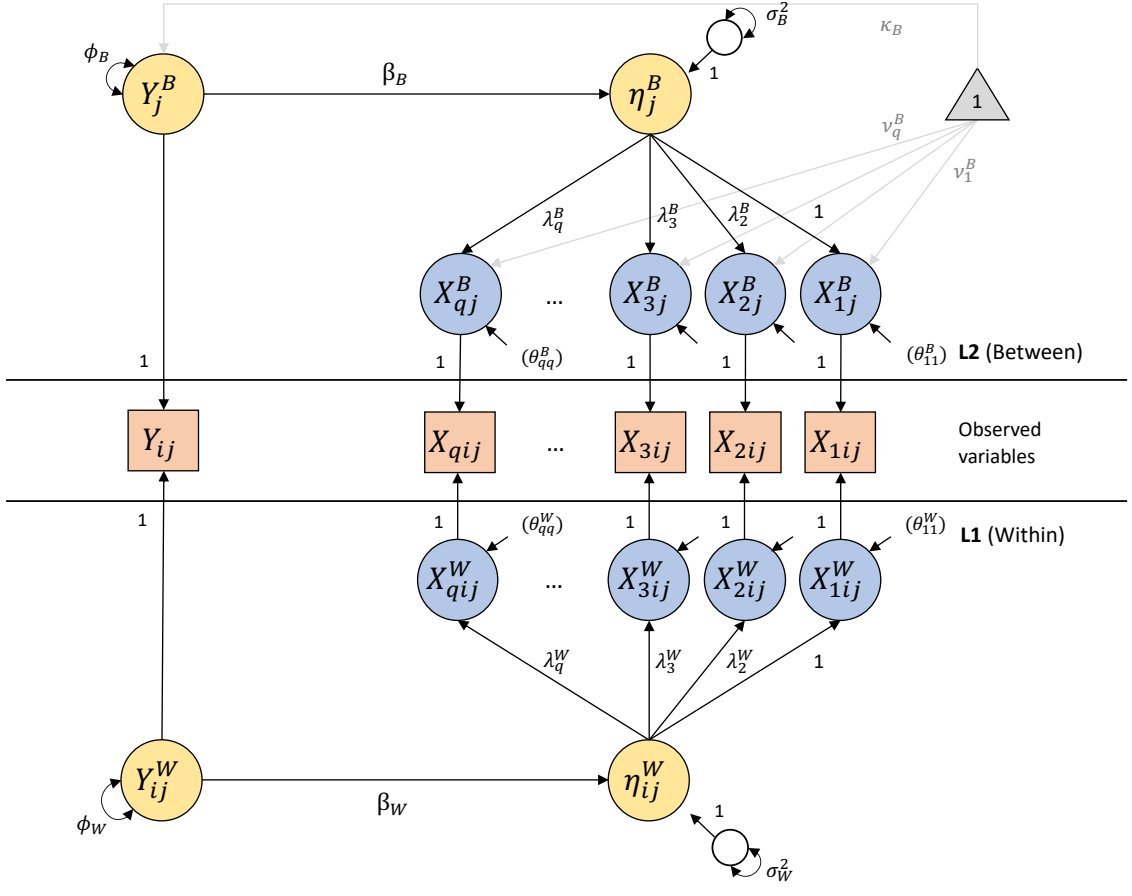


3. Data preparation

Constructs of interest	• Mathematics achievement • Self-concept
Complex data structure	• Study ID • Country ID • Classroom ID • Students ID • Student weights • Classroom weights
Country-level variables	• Socio-economic development • Cultural dimensions

4. Analysis of IPD and model-based effect size estimation

Multilevel SEM to estimate contextual effects



4. Analysis of IPD and model-based effect size estimation

Multi-Group Multilevel CFA: Measurement Invariance Testing across Levels and Countries

Model	RMSEA	CFI	SRMR-W	SRMR-B	$\Delta\chi^2(\Delta df)$
TIMSS 2003					
Configural model	.000	.998	.00	.04	
L1 invariance model	.077	.947	.04	.08	2829.83 (113)*
L2 invariance model	.048	.959	.04	.14	241.21 (113)*
L1 and L2 invariance model	.047	.958	.04	.12	72.18 (14)*
Contextual model	.053	.931	.04	.14	

5. Meta-analysis of the effect sizes

ESID	Year	Country	*N*	BFLPE	*var*	PDI	IDV	MAS	UAI	LTO	IVR	HDI
1	2003	AUS	4225	-0.60	0.009	38	90	61	51	21	71	0.91
2	2003	BFL	4668	-0.45	0.003	65	75	54	94	82	57	0.89
3	2003	COT	4284	-0.27	0.003	39	80	52	48	36	68	0.88
...	...	...	...	...	...	...	...	...	...	...	...	...
200	2019	AAD	8621	-0.62	0.007	80	38	53	68	23	34	0.89
201	2019	ADU	7251	-0.44	0.003	80	38	53	68	23	34	0.89
202	2019	ARE	25173	-0.62	0.002	80	38	53	68	23	34	0.89
203	2019	AUS	5764	-0.49	0.003	38	90	61	51	21.15	71.42	0.94
204	2019	AUT	4420	-0.45	0.003	11	55	79	70	60.45	62.72	0.92
...	...	...	...	...	...	...	...	...	...	...	...	...

5. Meta-analysis of the effect sizes

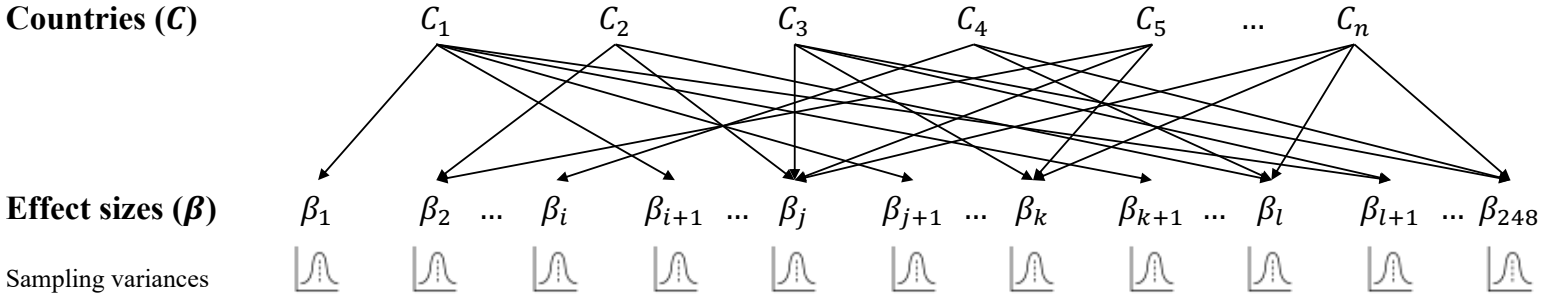
Three-level random-effects model with effect sizes nested in countries

Level 1 (sampling variance): $\beta_{jk} = \theta_{jk} + r_{jk}$

Level 2 (within countries): $\theta_{jk} = \lambda_k + q_{jk}$

Level 3 (between countries): $\lambda_k = \mu + u_k$

$r_{jk} \sim N(0, \sigma_r^2)$, $q_{jk} \sim N(0, \sigma_q^2)$, $u_k \sim N(0, \sigma_u^2)$



5. Meta-analysis of the effect sizes

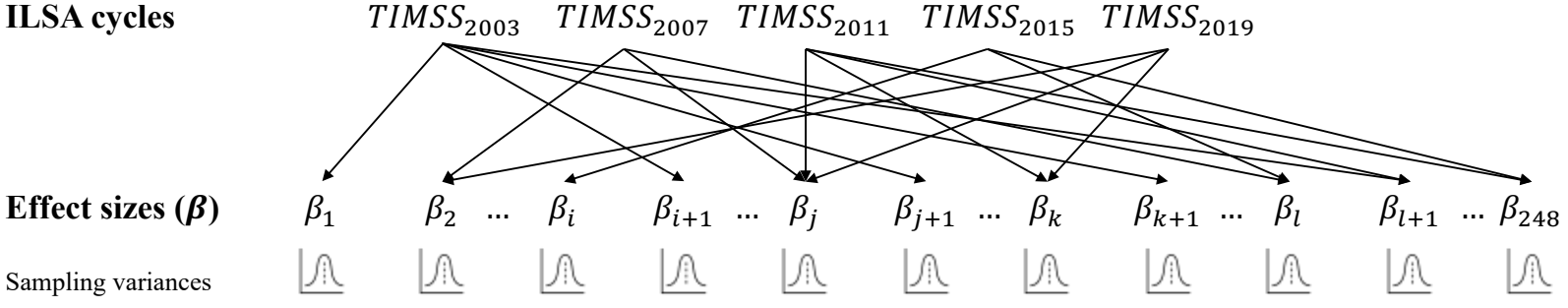
Three-level random-effects model with effect sizes nested in cycles

Level 1 (sampling variance): $\beta_{jk} = \theta_{jk} + r_{jk}$

Level 2 (within cycles): $\theta_{jk} = \lambda_k + q_{jk}$

Level 3 (between cycles): $\lambda_k = \mu + u_k$

$r_{jk} \sim N(0, \sigma_r^2)$, $q_{jk} \sim N(0, \sigma_q^2)$, $u_k \sim N(0, \sigma_u^2)$



5. Meta-analysis of the effect sizes

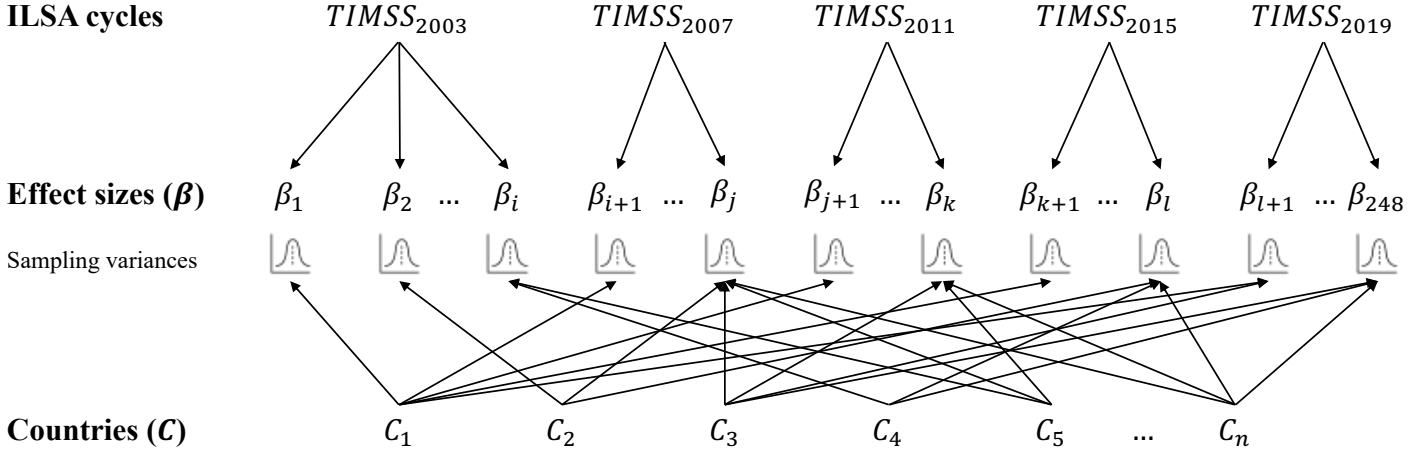
Four-level cross-classified random-effects model with countries and cycles

Level 1: $\beta_{j(kl)} = \theta_{j(kl)} + r_{j(kl)}$

Level 2: $\theta_{j(kl)} = \lambda_{(kl)} + q_{j(kl)}$

Levels 3 and 4: $\lambda_{(kl)} = \mu + u_k + p_l$

$r_{j(kl)} \sim N(0, \sigma_r^2)$, $q_{j(kl)} \sim N(0, \sigma_q^2)$,
$u_k \sim N(0, \sigma_u^2)$, $p_l \sim N(0, \sigma_p^2)$



5. Meta-analysis of the effect sizes

Baseline model	$\hat{\mu}$ [95 % CI]	σ^2_{ES} [95 % CI]	σ^2_C [95 % CI]	σ^2_Y [95 % CI]	I^2_{ES}	I^2_C	I^2_Y
Model 1: Standard random-effects model	-.458 [-.481, -.436]	.025 [.020, .030]	-	-	82.4 %	-	-
Model 2: Three-level random-effects model with effect sizes nested in countries	-.451 [-.486, -.416]	.004 [.002, .006]	.021 [.014, .031]	-	12.1 %	70.0 %	-
Model 3: Three-level random-effects model with effect sizes nested in TIMSS cycles	-.458 [-.484, -.432]	.025 [.020, .031]	-	.000 [.000, .002]	82.4 %	-	0.0 %
Model 5: Four-level cross-classified random-effects model	-.452 [-.489, -.415]	.003 [.001, .006]	.021 [.014, .031]	.000 [.000, .002]	10.9 %	70.6 %	0.7 %

5. Meta-analysis of the effect sizes

Moderator	Three-level mixed-effects meta-regression			
	Model 2a		Model 2b	
	B	*SE*	*B*	*SE*
Intercept	-0.137	0.174	-0.812	0.225
Cultural dimensions				
PDI	-0.002	0.002	-	-
IDV	-0.002	0.001	-	-
MAS	-0.001	0.001	-	-
UAI	-0.001	0.001	-	-
LTO	0.000	0.001	-	-
IVR	0.000	0.002	-	-
Economic development			-	-
HDI	-	-	0.426*	0.258
Moderator test				
Q_M(df)	8.6 (6), p = .20		5.5 (1), p = .02	
Variance explanation				
R^2_{ES}	0.0%		2.5%	
R^2_C	5.5%		12.3%	

6. Reporting and documenting

- Interpretation and communication of meta-analytic findings
- Publications of reports, analytic code, and data (open science for replicability)

Potential of Two-Stage IPD Meta-Analysis

1. Two-stage IPD meta-analysis enables researchers to synthesize information from complex surveys studies

2. Standardized analyses across studies

3. Direct and model-based generation of the effect sizes of interest

4. Appropriate handling of statistical dependencies in meta-analytic data sets from complex sampling surveys.

UiO : CEMO – Centre for Educational Measurement
University of Oslo

A Primer on Two-Stage Meta-Analysis With Individual Participant Data Obtained From Complex Sampling Surveys

Diego G. Campos[1], Mike W.-L. Cheung[2], and Ronny Scherer[1]

[1] Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo

[2] Department of Psychology, National University of Singapore

Thank you!

Questions?

Contact
Email: diegogc@uio.no
Twitter: @diegc_

References

- Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/metaanalyze approach. *Frontiers in Psychology, 7(MAY)*, 1–13. https://doi.org/10.3389/fpsyg.2016.00738
- Holzberger, D., Reinhold, S., Lüdtke, O., & Seidel, T. (2020). A meta-analysis on the relationship between school characteristics and student outcomes in science and maths–evidence from large-scale studies. *Studies in Science Education*, *56*(1), 1-34
- Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). A demonstration and evaluation of the use of crossclassifiedrandom-effects models for meta-analysis. *Behavior Research Methods, 51*(3), 1286–1304. https://doi.org/10.3758/s13428-018-1063-2
- Jak, S., & Cheung, M. W.-L. (2019). Meta-analytic structural equation modeling with moderating effects on SEM parameters. Psychological Methods. https://doi.org/10.1037/met0000245
- Keller, L., Preckel, F., & Brunner, M. (2021). Nonlinear relations between achievement and academic self-concepts in elementary and secondary school: An integrative data analysis across 13 countries. *Journal of Educational Psychology, 113*(3), 585–604. https://doi-org.ezproxy.uio.no/10.1037/edu0000533
- Mittal O., Nilsen T., Björnsson J.K. (2020) Measuring Equity Across the Nordic Education Systems—Conceptual and Methodological Choices as Implications for Educational Policies. In: Frønes T.S., Pettersen A., Radišić J., Buchholtz N. (eds) Equity, Equality and Diversity in the Nordic Model of Education. Springer, Cham. https://doi-org.ezproxy.uio.no/10.1007/978-3-030-61648-9_3
- Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053. https://doi.org/10.1037/a0027697
- OECD. (2017). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. OECD. https://doi.org/10.1787/9789264281820-en
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher, 39*(2), 142–151. https://doi.org/10.3102/0013189X10363170

5. Meta-analysis of the effect sizes

- The three-level cycles model had a better fit than the cross-classified random effects model, $\chi^2(3) = 117.843, p < 0.0001$
- The cross-classified random effects model and the three-level countries model showed similar fit, $\chi^2(3) = 1.665, p = 0.197$.
- There was no gain when adding the between-cycles variation.