UiO : CEMO – Centre for Educational Measurement
University of Oslo

Meta-Analyzing International Large-Scale Assessment Data: An Application of the Split, Analyze, and Meta-Analyze (SAM) Approach

Diego Campos*; Ronny Scherer*; Mike Cheung**

*Centre For Educational Measurement (CEMO), Faculty of Educational Sciences, University of Oslo, Norway
**Department of Psychology, National University of Singapore, Singapore, Singapore

Background

Complex survey studies, such as International Large Scale Assessments (ILSAs), provide large-scale data of students, classrooms, and schools that can be used to:

- Test hypotheses about relations among constructs
- Study differences between groups
- Attempt large-scale replications

Photo by Hello I'm Nik on Unsplash

OECD, (2017)

Background

Synthesis of ILSA Data

- Are the effect sizes from ILSA data consistent across countries, waves, ILSAs?
- What is the average effect size of a relationship between two variables?
- Are group/individual characteristics related to the variability in the effect sizes?

Background

Challenges in secondary analyses and meta-analyses

1. Account for the hierarchical structure of the data
   - Primary nesting
   - Secondary nesting
2. Incorporate sample weights
3. Plausible values
4. Comparability across countries

Photo by Volodymyr Hryshchenko on Unsplash

Rutkowski, Gonzalez, Joncas, von Davier (2010)

Background

- ILSA studies have not been considered in most meta-analyses

- Methodological choices in the analysis of ILSA data lead to substantially different results

Photo by Markus Spiske on Unsplash

Mittal, Nilsen, & Björnsson (2020)

Three steps to meta-analyse ILSA data

Step 1 – Split

ILSA

$Wave_1$ $Wave_2$ $Wave_3$ $Wave_k$

$Country_k$ $Country_k$ $Country_k$ $Country_k$

Step 2 – Analyse

$\boldsymbol{\beta_k}$

L2 X_{JK} Y_{JK}

L1 X_{ijk} Y_{ijk}

Step 3 – Meta-analyse

$$y_i = \beta_0 + u_i + e_i$$

Cheung & Jack (2016)

Illustrative Case

We aimed to study the generalizability of the Big-Fish-Little-Pond-Effect (BFLPE) across countries and TIMSS waves

- RQ1: To what extent does evidence for the generalizability of the BFLPE and its variation across countries exist?
- RQ2: To what extent do country-level variables (Human development index, individualism, power distance, uncertainty avoidance, masculinity, indulgence, and long-term orientation) explain the variation in the BFLPE across countries?

Method

Five TIMSS cycles 2003, 2007, 2011, 2015, and 2019

- All the participating countries
- Fourth-grade data only

Table 1

Sample size

	Countries	Sample Size
ILSA Cycle		
TIMSS 2003	29	123815
TIMSS 2007	44	182488
TIMSS 2011	55	277493
TIMSS 2015	56	323299
TIMSS 2019	64	330191
Meta-analysis		
Model	**85**	**248**

UiO CEMO – Centre for Educational Measurement

Data Analysis

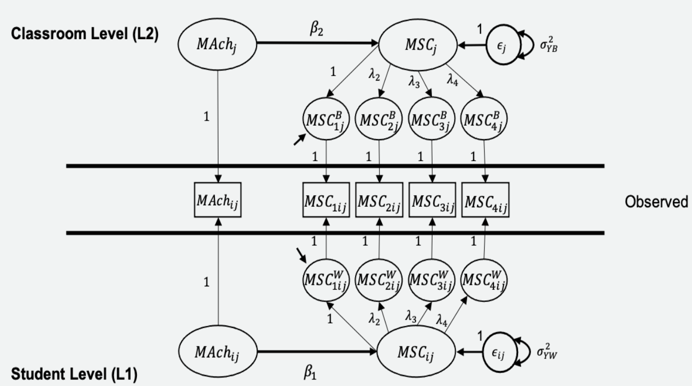
Step 1 Split

TIMSS

$TIMSS_{2003}$ $TIMSS_{2007}$ $TIMSS_{2011}$ $TIMSS_k$

$Country_k$ $Country_k$ $Country_k$ $Country_k$

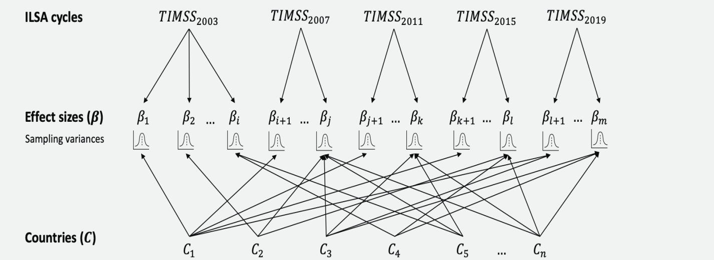
Step 2 Analyse



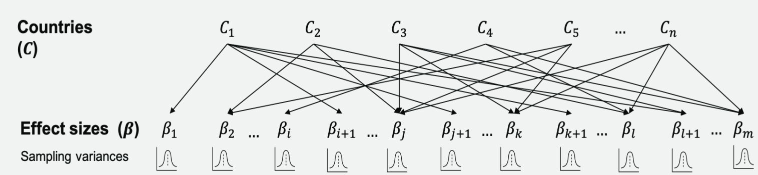
$$ES_{BFLPE} = 2 \times (\beta_j - \beta_i) \times \sigma_{yj} / \sigma_{yij}$$

Step 3 Meta-analyse

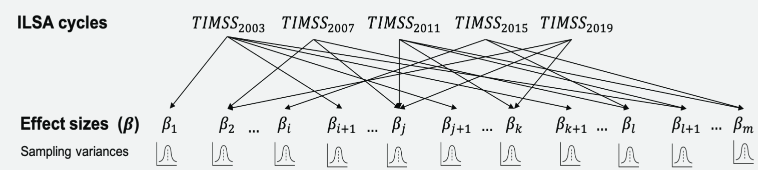
Cross-classified random-effects model



Three-level model - Countries



Three-level model – Cycles



Results Meta-Analysis – Comparison working models

- Ignoring the nesting merge the variance associated to variation in the effect sizes and countries
- TIMSS nesting does not seem to explain variance in the effect sizes
- The estimated average effect size is similar across the working models

Table 3

Comparison of meta-analytic models

	$\hat{\mu}$ [CI]	$\sigma^2_{(1)}$	$\sigma^2_{(2)}$	$\sigma^2_{(3)}$	I^2	$I^2_{(2)}$	$I^2_{(3)}$
Random-effects model	-.458 [-.480, -.435]	.025 [.02, .03]			82.43%	-	-
Three-level Waves	-.458 [-.484, -.432]	-	.025 [.020, .031]	.000 [.000, .002]	-	82.43%	0%
Three-level Countries	-.451 [-.486, -.416]	-	.004 [.002, .006]	.021 [.014, .031]	-	12.06%	69.97%
Cross-classified	-.452 [-.489, -.415]	.020 [.014, .031]	.003 [.001, .006]	.000 [.000, .002]	70.58%	10.87%	0.69%

Results Meta-Analysis – Comparison working models

- The three-level waves 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 is no gain when adding the between-waves variation

Results - Three-level mixed-effects meta-regression

The cultural characteristics did not explain the heterogeneity in the effect sizes

Table 4

Results multivariate mixed-effects meta-analysis - cultural variables

	Slope Coefficient	95% CI	p-value
Intercept	-.137	[-.561, .140]	.440
Individualism - Collectivism	-.002	[-.005, -.0002]	.142
Masculinity-femininity	-.000	[-.003, .002]	.492
Power distance	-.002	[-.006, .001]	.148
Uncertainty avoidance	-.001	[-.002, .001]	.276
Indulgence-restraint	.002	[-.001, .005]	.823
Long-term vs. short-term orientation	.000	[-.002, .002]	.795

Results - Three-level mixed-effects meta-regression

The economic index did not explain the heterogeneity in the effect sizes

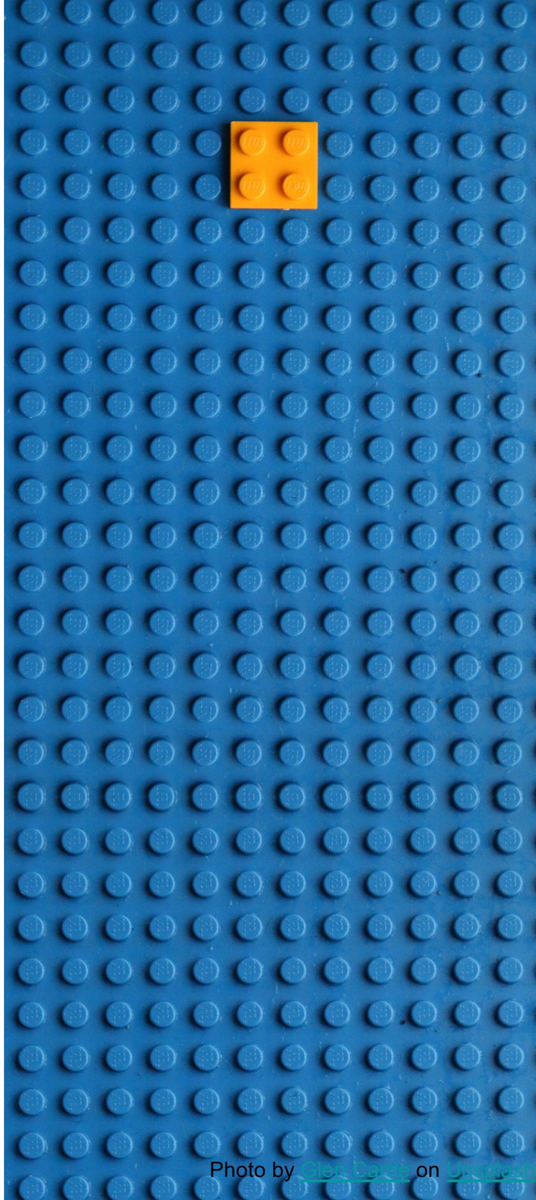
Table 5

Results multivariate mixed-effects meta-analysis economic index

	Slope Coefficient	95% CI	p-value
Intercept	-.812	[-1.114, -.509]	<.0001
Human Development Index	.426	[.069, .782]	.107

Conclusions

1. Cross-country differences and variation between effect sizes explain most of the heterogeneity in the BFLPE
2. Ignoring the nesting of the meta-analytic dataset conflate the variance associated with the effect sizes and the countries
3. ILSA cycles do not account for variance in the effect sizes



Conclusions

4. None of the country-level variables explain the heterogeneity between effect sizes
5. The three-level model with nesting in countries had the best performance among all of the working models

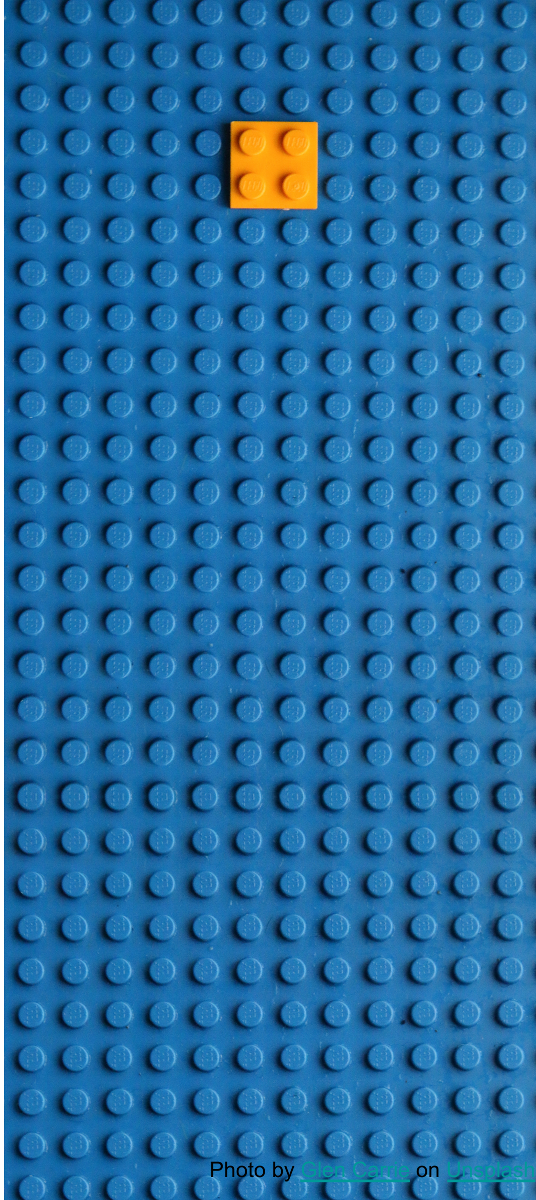


Photo by Glen Carrie on Unsplash

Implications

1. The meta-analysis of ILSA studies has multiple benefits:
   - Derive multiple effect sizes
   - Test for the comparability of measures across groups
   - Test for the replicability of findings in high-quality data with representative samples
2. The SAM approach can be used to synthesize effect sizes from complex survey data
3. The meta-analysis of ILSA data must consider the characteristics of the primary dataset and the nested structure of the meta-analytic dataset

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
- 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
- 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

Thank you!

Questions?