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

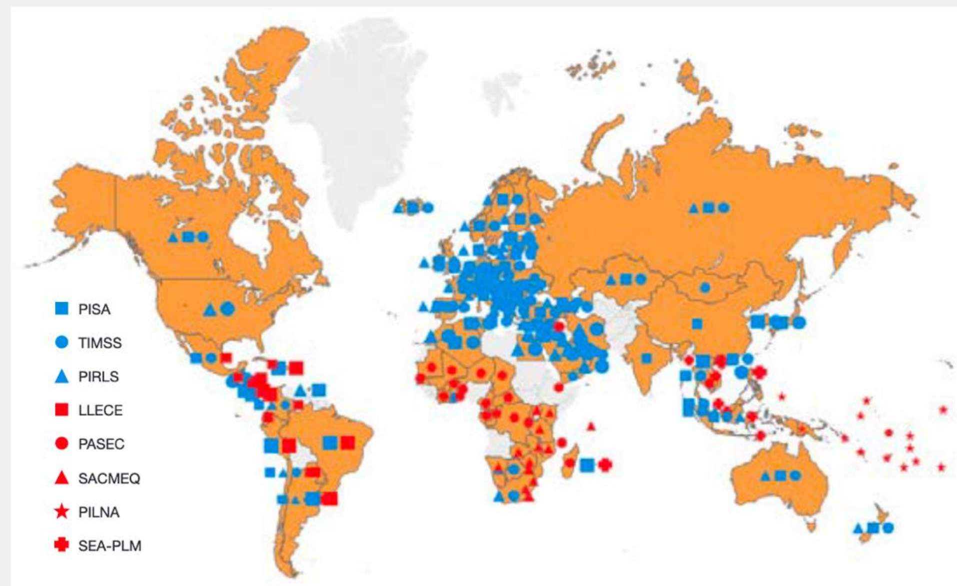
Diego Campos*; Ronny Scherer*; Mike Cheung**

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

# Background

Regional, national, and international educational assessments are increasing around the world

**Figure 2.2 School-based assessment**




**Note:** Areas shaded in orange correspond to the existence of national assessments.
*Source: UNESCO Institute for Statistics (UIS).*

UNESCO, (2018)

# Background

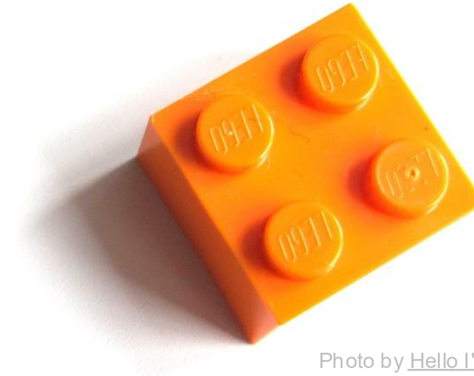


Photo by Hello I'm Nik on Unsplash

Large-scale educational assessments provide 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

## Advantages

- Representative probability samples
- Standardized measurements

OECD, (2017)

# Background

- The assessments have not been considered in most meta-analyses

  Holzberger, Reinhold, Lüdtke, Seidel (2020)
  Keller, L., Preckel, F., & Brunner, M. (2021)

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

  Mittal, Nilsen, & Björnsson (2020)

Photo by Markus Spiske on Unsplash

# Background

## Challenges for the meta-analyses

1. Cluster sampling
   - Primary nesting
   - Secondary nesting
2. Unequal selection probability
3. Plausible values
4. Comparability across countries

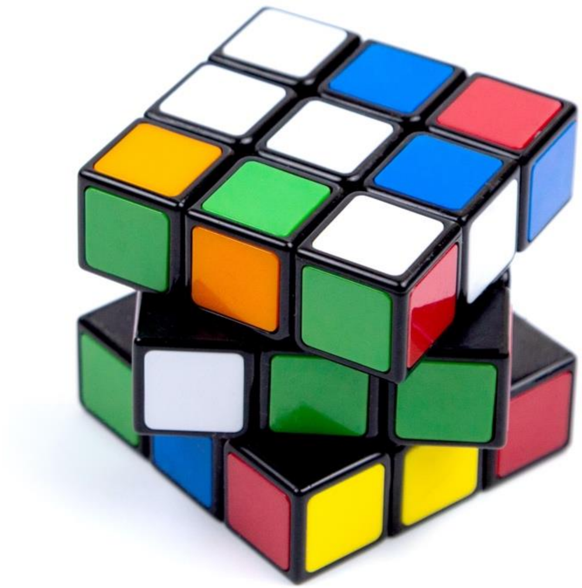


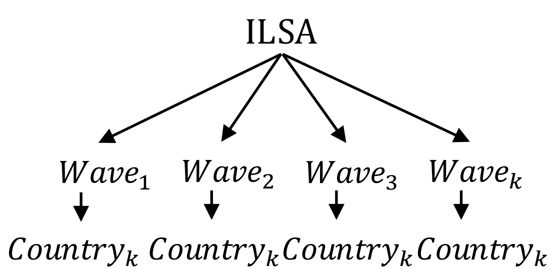
Photo by Volodymyr Hryshchenko on Unsplash

Rutkowski, Gonzalez, Joncas, von Davier (2010)

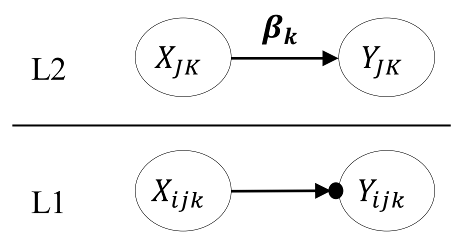
# Three steps to meta-analyse ILSA data

## Individual Participant Data Meta-analysis

### Step 1 – Split




### Step 2 – Analyse




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

### 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 support the generalizability of the BFLPE across countries?
- 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

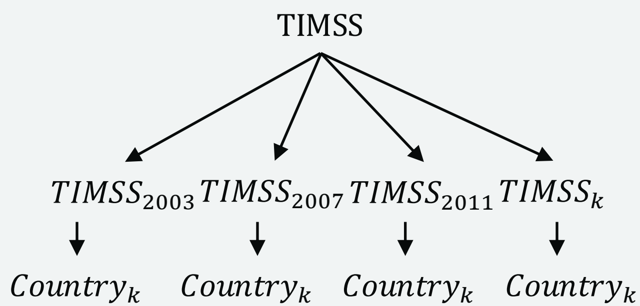
- All the participating countries
- Fourth-grade data only

**Table 1**
*Sample size*

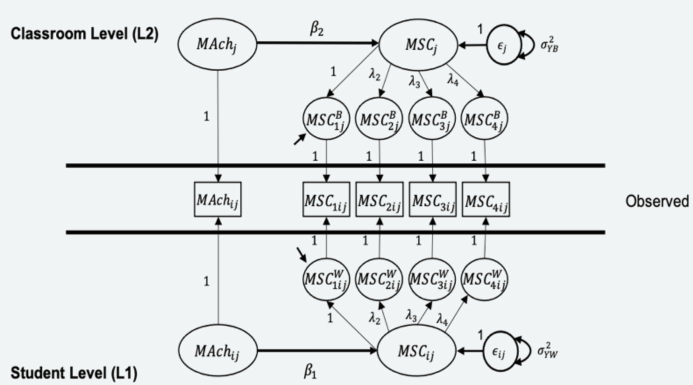
| | Countries | Sample Size |
|---|---|---|
| ILSA Cycle | | |
| TIMSS 2003 | 29 | 123815 |
| TIMSS 2007 | 44 | 182488 |
| TIMSS 2011 | 55 | 277493 |
| TIMSS 2015 | 56 | 323299 |
| TIMSS 2019 | 64 | 330191 |
| Meta-analysis | | |
| Model | **85** | **248** |

# Data Analysis

**Step 1 Split**



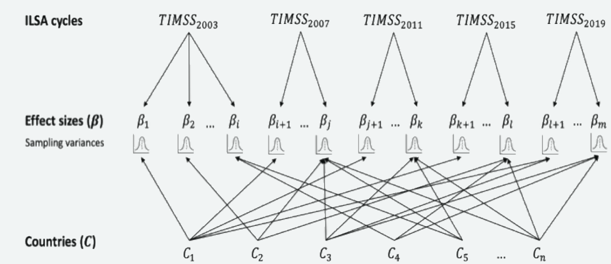

**Step 2 Analyse**



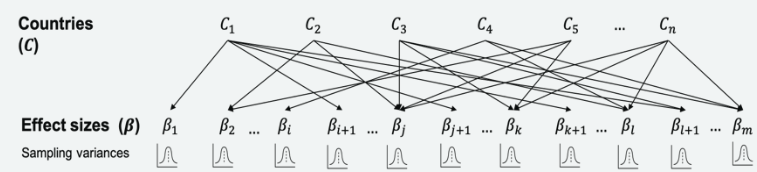

$$ES_{BFLPE} = 2 \times (\beta_j - \beta_i) \times \sigma_{yj} / \sigma_{yij}$$

**Step 3 Meta-analyse**

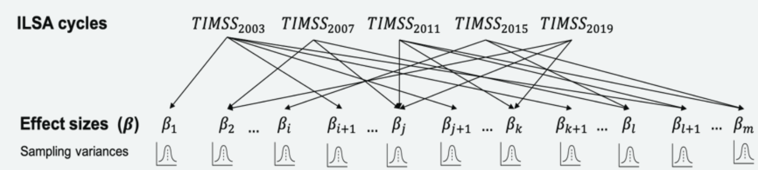
## Cross-classified random-effects model




## Three-level model - Countries




## Three-level model – Cycles



# Results Meta-Analysis – Comparison working models

**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% |

- The estimated average effect size is similar across the 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

# 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

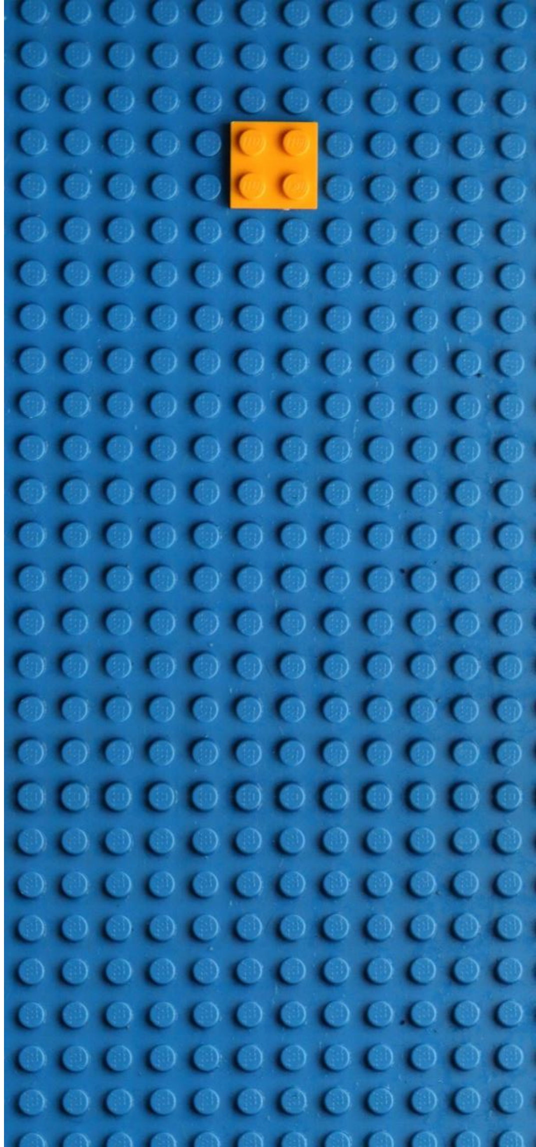


Photo by Glen Carrie on Unsplash

# Conclusions

4. None of the country-level variables explain the heterogeneity between effect sizes

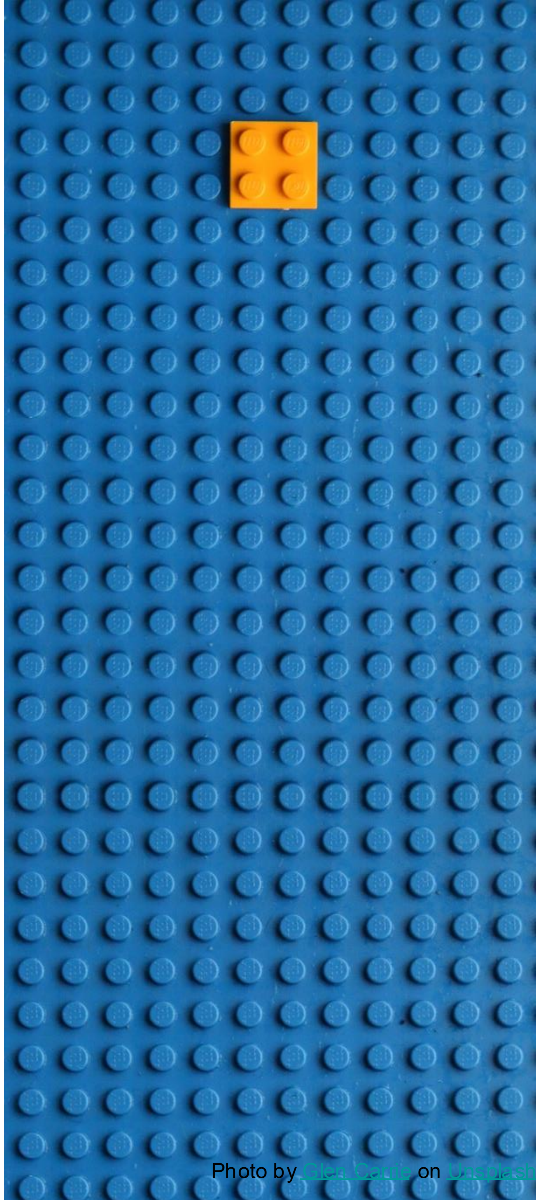


Photo by Glen Carrie on Unsplash

# 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 methodological complexities of the ILSA 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
- 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

# Thank you!

# **Questions?**

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

**Table 2.**

*Description Cultural Dimensions proposed by Hofstede et al. (2010)*

| Dimension | Description |
|---|---|
| Individualism-Collectivism | Refers to the extent to which people feel independent, rather than interdependent as members of larger groups. |
| Power distance | Denotes the degree to which nationals from a country accept and expect that power is distributed unequally. |
| Uncertainty avoidance | Indicates the degree to which members of a society are tolerant to uncertainty and ambiguity. |
| Masculinity-feminity | Specifies the extent to which the use of force is socially approved within a culture. |
| Long-term vs. short-term orientation | Reveals the degree to which a culture is focused on pragmatic solutions and adaptations to future challenges versus traditional solutions based on the past and traditions. |
| Indulgence-restraint | Identifies the degree to which a society allows or supress the gratification of needs related to enjoying life and having fun. |

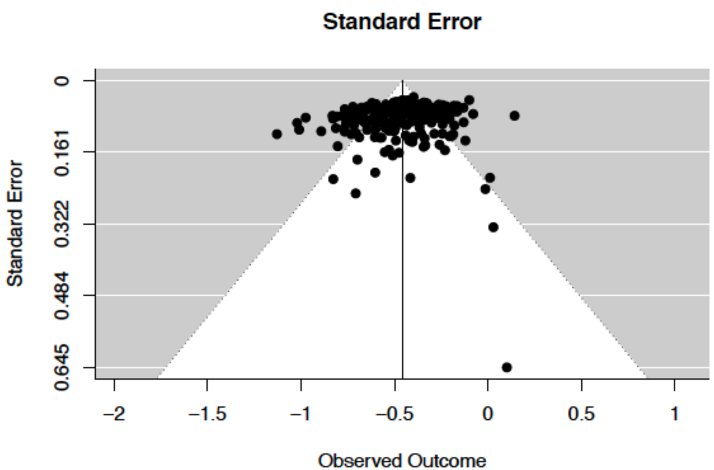
# Results – MG MCFA

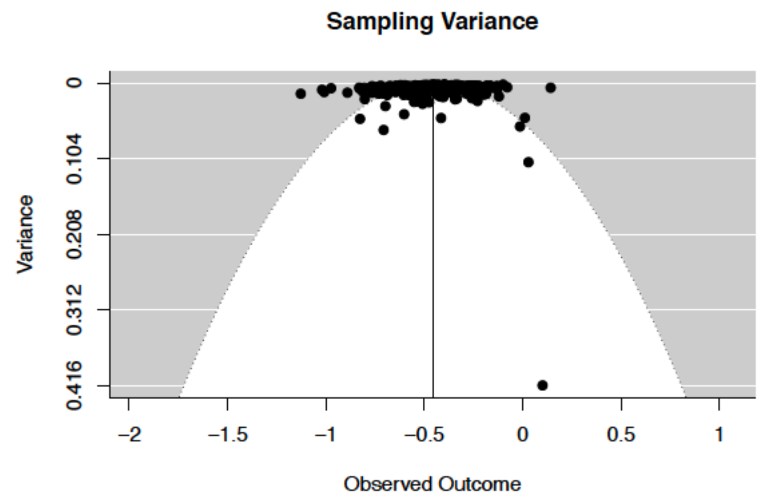
**Table 1**

*Fit statistics of the measurement invariance models (MG-MCFA)*

| | TIMSS 2003 | | | TIMSS 2007 | | | TIMSS 2011 | | | TIMSS 2015 | | | TIMSS 2019 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | RMSEA | CFI | TLI | RMSEA | CFI | TLI | RMSEA | CFI | TLI | RMSEA | CFI | TLI | RMSEA | CFI | TLI |
| Configural model | .000 | .998 | 1.000 | .000 | .995 | 1.000 | .000 | .996 | 1.000 | .027 | 1.000 | .98 | .027 | 1.000 | .982 |
| L1 invariance model | .077 | .947 | .836 | .066 | .956 | .867 | .061 | .968 | .902 | .063 | .963 | .889 | .067 | .962 | .886 |
| L2 invariance model | .048 | .959 | .937 | .040 | .969 | .952 | .037 | .977 | .965 | .036 | .976 | .963 | .04 | .973 | .96 |
| L1 and L2 invariance model | .047 | .958 | .939 | .039 | .969 | .954 | .036 | .977 | .966 | .036 | .975 | .965 | .039 | .973 | .962 |
| Contextual model | .053 | .931 | .904 | .051 | .934 | .908 | .056 | .927 | .897 | .061 | .910 | .875 | .057 | .924 | .896 |

*Note.* Configural model: a model with one factor for mathematics self-concept and factor loadings allowed to vary across levels and countries. L1 invariance model: factor loadings restricted to be invariant across countries at L1. L2 invariance model: factor loadings are restricted to be invariant across countries at L1 and L2, but with factor loadings allowed to vary across levels. L1 and L2 Invariance Model: a model with factor loadings restricted to be invariant across countries and levels. The contextual model of TIMSS 2011 is currently under analysis.


Standard Error
Standard Error
0
0.161
0.322
0.484
0.645
−2
−1.5
−1
−0.5
0
0.5
1
Observed Outcome


Sampling Variance
Variance
0
0.104
0.208
0.312
0.416
−2
−1.5
−1
−0.5
0
0.5
1
Observed Outcome

UiO CEMO – Centre for Educational Measurement
University of Oslo