The factor structure of mathematical abilities in Luxembourg's national school monitoring: Its stability over elementary school and relations to, gender, language background, and SES

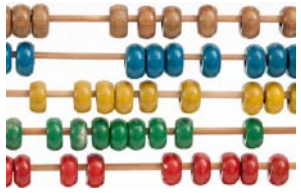
Dr. Philipp Sonnleitner & Dr. Caroline Hornung

LUCET | LUXEMBOURG CENTRE FOR EDUCATIONAL TESTING

UNIVERSITÉ DU LUXEMBOURG

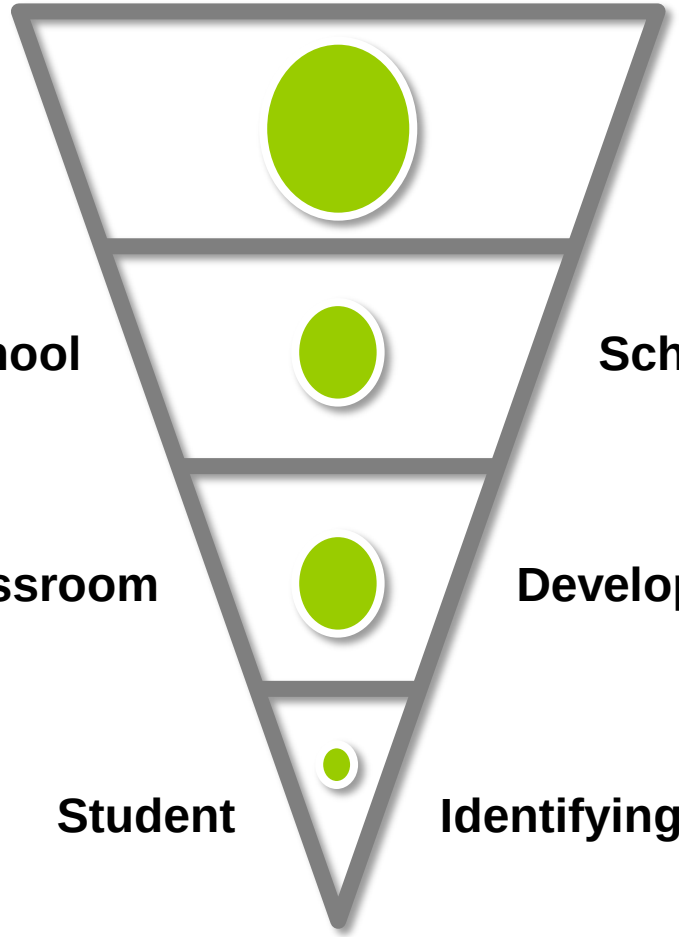
Aim of the study

- **Mathematics** is the fundament of modern societies and the starting point for all STEM-related fields
- Math skills are key **predictors of academic success** (e.g. Lyons & Ansari, 2015)
- At the center of educational **Large-Scale Assessments**
- Math skills are **multidimensional** (e.g., Bräuning et al. 2020; Clements, et al. 2008; Gnaldi, 2017; Milburn et al. 2019; Saβ et al. 2017)
- But psychometric analyses often focus on **one general mathematical ability** (single latent factor) (Saβ et al. 2017)
- How and to what extent **does this simplification affect educational studies** that rely on these data?



The Luxembourgish national school monitoring (ÉpStan) has numerous aims



System — **Development of school quality and evaluation**

National Report and Educational Report

School — **School development**

Schoolreport

Classroom — **Development of teaching practice**

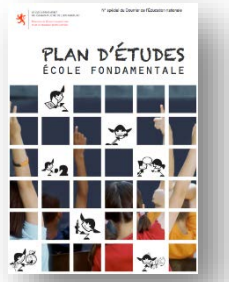
Classroom Report (for teachers)

Student — **Identifying special educational needs**

Student Feedback (for teachers, students and parents)

Test framework: Mathematical dimensions assessed throughout the grades 1-5 (and 7)

- The content of the math tests build on the **national curriculum**
  - "Number & Operations" and "Shape & Space"
- Different **difficulty levels**

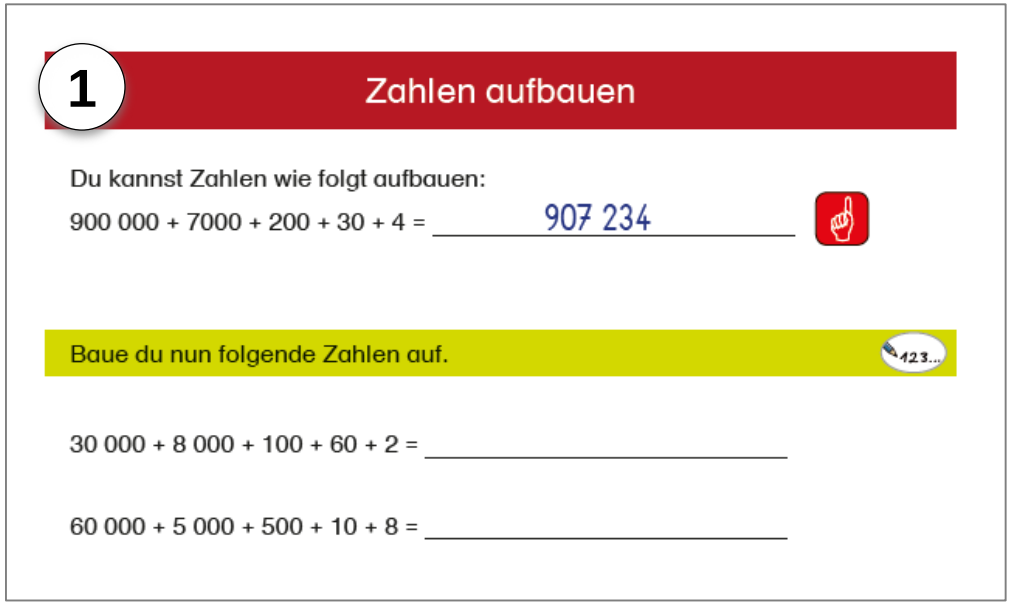


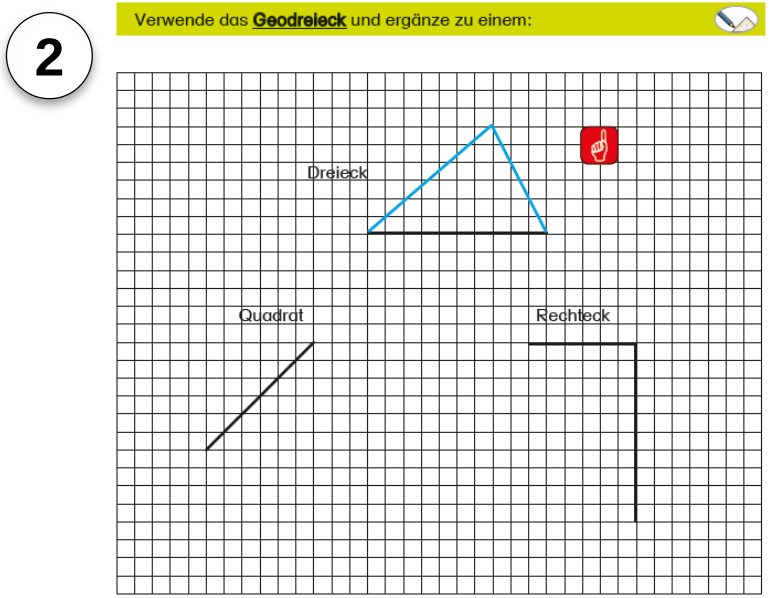
Mathematics

ÉpStan	Plan d'Études Enseignement Fondamental									
	Level 1	**Level 2**	**Level 3**	**Level 4**	**Level 5**	**Level 6**	**Level 7**	**Level 8**	**Level 9**	**Level 10**
	kindergarten	Grade 1	Grade 2							
Grade 1	+++	++	+							
Grade 3	+	+	+++	++	+					
Grade 5			+	+	+++	++	+	+		
Grade 7					+	+	+	+++	++	+

+++ approximately 50% of test items

Test framework: item examples : Grade 5





Both items measure „difficulty level 5“

Item 1 measures „number and operations“
Item 2 measures „space and shape“

Research questions

- Can we find the two mathematical dimensions Space & Shape and Number & Operations also in our data?
- Is the latent structure of the mathematics abilities as assessed by our national school monitoring program stable over elementary school, ranging from Grade 1 to Grade 5?
- Does it matter how we conceptualize math ability when studying the relation to known relevant variables, such as gender, language background or socioeconomic status?

Sample and Test Description

- Full cohorts for G1-3-5 of 2019

	N	Mean age in years (SD)	Girls
Grade 1	5807	6.4 (.52)	2845 (49.9%)
Grade 3	5456	8.5 (.62)	2639 (49%)
Grade 5	5200	10.6 (.68)	2565 (49.8%)

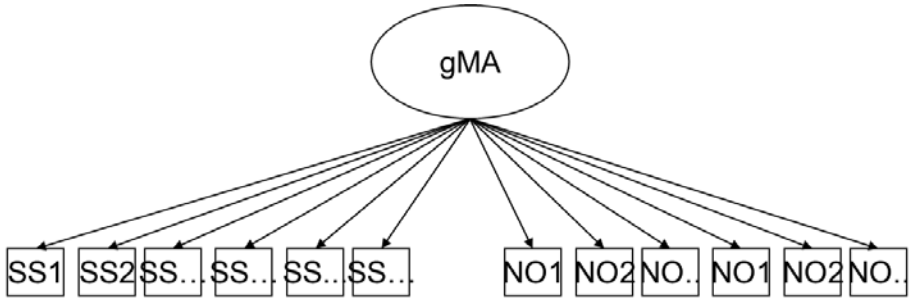
- Test Booklets of 2019

	n items	Space & Shape	Numbers & Operation
Grade 1	48	16	32
Grade 3	69	27	42
Grade 5	60	23	37

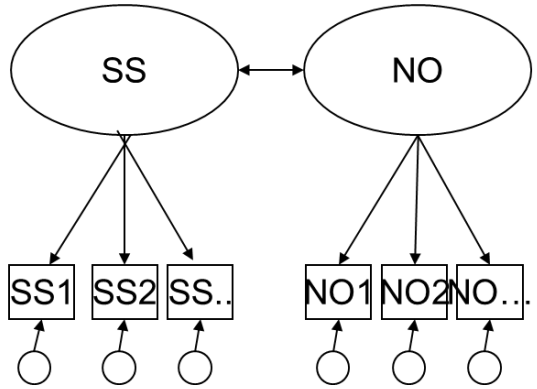
➔ Items typically scaled using a 1-PL model and screened for infit (>.80 & <1.20), rit (>.25), Cohort DIF

Confirmatory factor analysis

- Two SEM are tested
  - **Single latent math construct** (1FM)



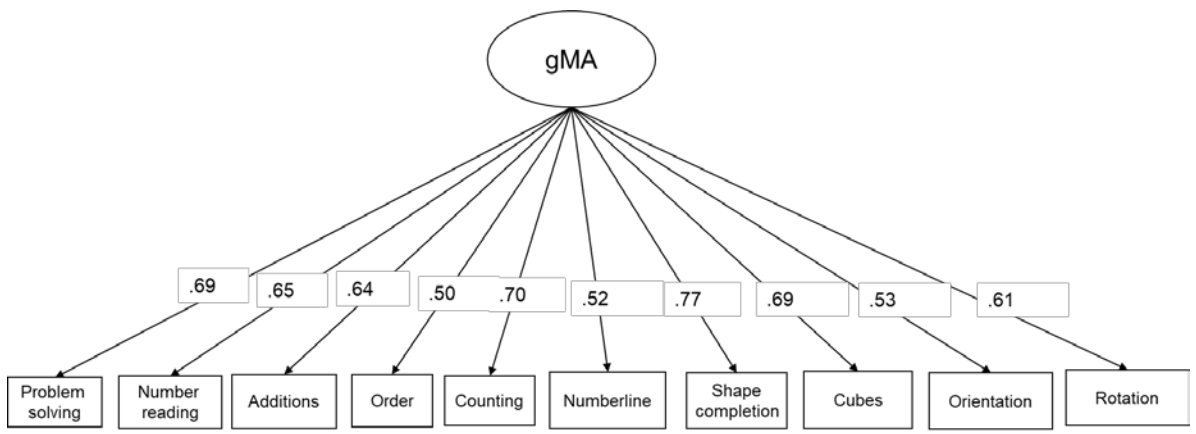
  - **Two related latent constructs:** (2FM)
    Space & Shape (SS) and Number & Operations (NO)

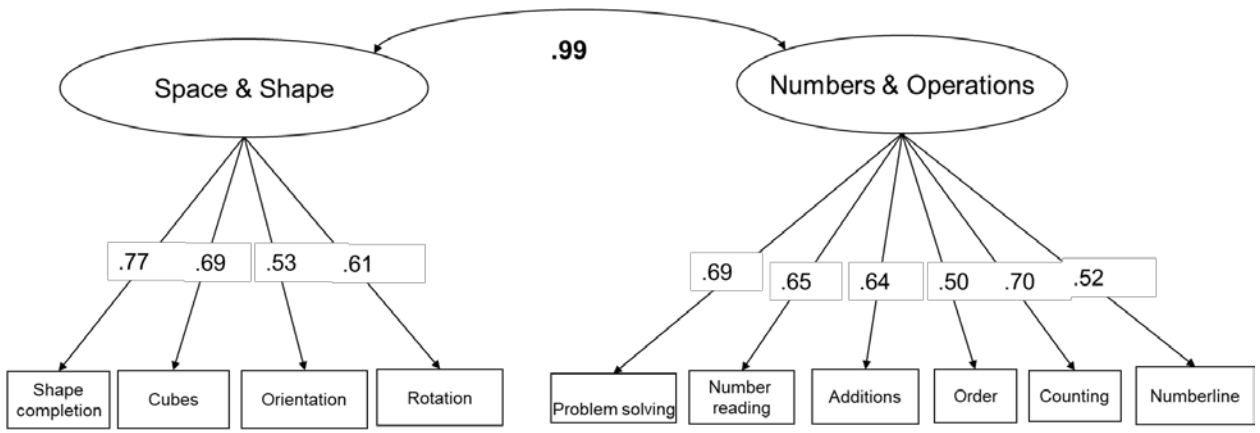


  - Parameter were estimated with Mplus 8.4 (Muthén & Muthén, 1998-2019)
  - homogenous item parcels captured the dimensional structure underlying the items (see Hall, Snell, & Singer Foust,1999)
  - Model fit evaluated according to the following fit indices (cut-off criteria in parentheses, see Fan & Sivo, 2007; Hu & Bentler, 1998):
    - CFI (> .95)
    - SRMR (< .08)
    - RMSEA (< .05)

Results CFA Grade 1

	Chi²	df	p	CFI	SRMR	RMSEA
1 Factor model	470.66	35	< .01	.98	.02	.05
2 Factor model	470.58	34	< .01	.98	.02	.05

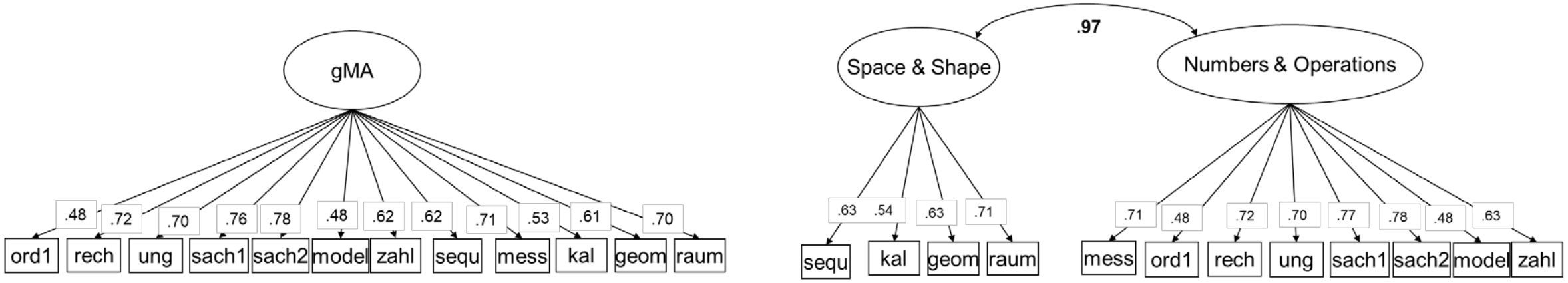




- Very good fit of both models
- Space & Shape not distinguishable from Numbers & Operations

Results CFA Grade 3

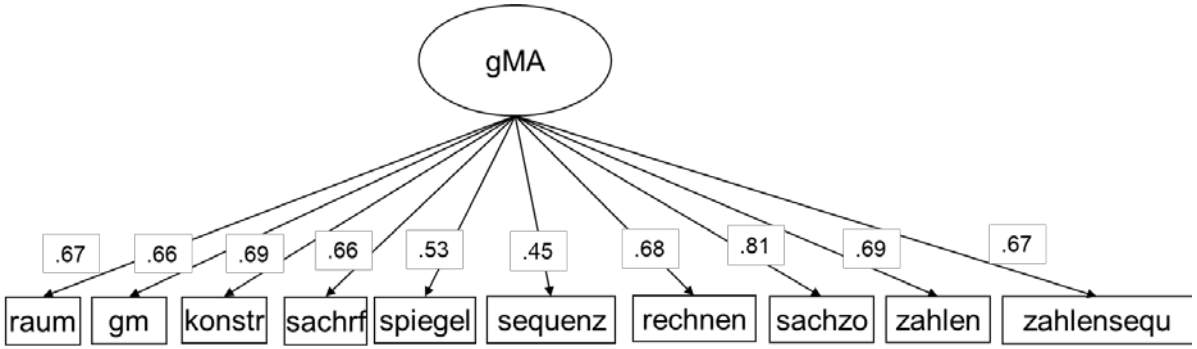
	Chi²	df	p	CFI	SRMR	RMSEA
1 Factor model	772.17	54	< .01	.97	.02	.05
2 Factor model	735.37	53	< .01	.97	.02	.05

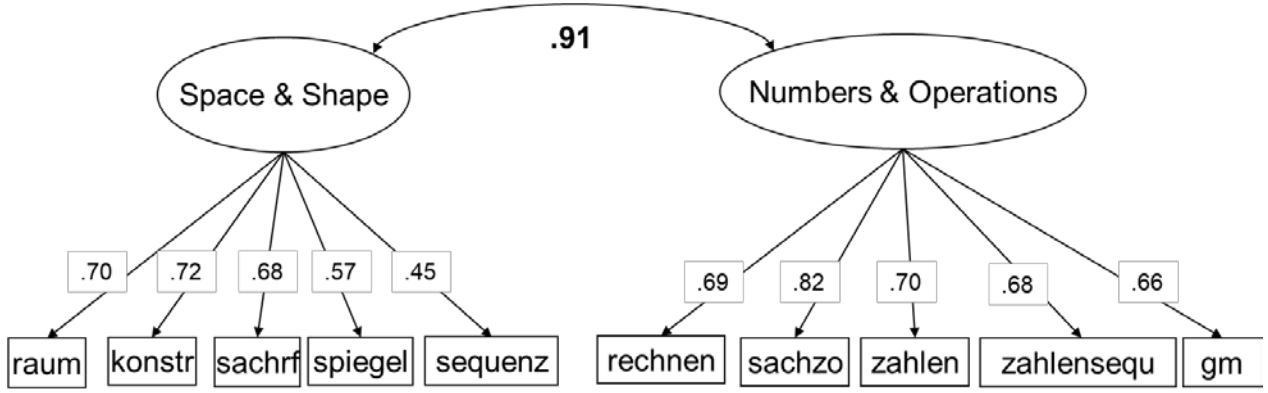


- (Again) Very good fit of both models
- Space & Shape empirically distinguishable from Numbers & Operations but almost identical

Results CFA Grade 5

	Chi²	df	p	CFI	SRMR	RMSEA
1 Factor model	697,67	35	< .01	.97	.03	.06
2 Factor model	426,52	34	< .01	.98	.02	.05





- Good fit of both models BUT!
- Space & Shape empirically distinguishable and differentiates from Numbers & Operations

Relations between math abilities and subject-specific variables in G5

	Model 1F: one factor	**Model 2F:** two factors	
Criterion	Math	Space & Shape	Number & Operations
Mathematics related constructs			
anxiety	**-.23**	**-.20**	**-.25**
interest	**.19**	**.13**	**.22**
self-concept	**.47**	**.37**	**.51**

- Structure of mathematical construct captured by ÉpStan makes a difference
- Math-related personality variables are more strongly linked to Number & Operations
- Number & Operations is cognitively overrepresented in associations with mathematics

Relationship between math abilities and sociodemographic variables in G5

		Model 1F: one factor	**Model 2F:** two factors	
Criterion		Math	Space & Shape	Number & Operations
		Latent correlation coefficients		
Socioeconomic status (HISEI)		**.37**	**.38**	**.36**
		Latent regression coefficients		
Gender	(1 = boys)	**.12**	.02	**.17**
Migration background	(1 = yes)	**-.14**	**-.16**	**-.13**
Home language	(1 = Lux./Ger.)	**.17**	**.19**	**.15**

- General math factor masks differential relations to sociodemographic variables
- Unexpected relation between gender and math subdomains
- Language and migration background more strongly influence Space & Shape domain

Summary & Outlook

- Math abilities captured by ÉpStan differentiate with increasing age despite unidimensional test development
- Math competencies become more specific over time (Deary et al., 1996)
- Results hint at importance of spatial language and vocabulary in Space & Shape items (Georges, Cornu, & Schiltz, 2020; Hornung et al., 2014)
- Mathematics subcompetencies should be considered (analyses, feedbacks, etc.)
- Multidimensional IRT-models for scaling might be promising

Feedback & Contact

philipp.sonnleitner@uni.lu

caroline.hornung@ext.uni.lu

Home of our research: https://wwwen.uni.lu/research/fhse/lucet

Test description and item examples: https://epstan.lu

More on our work: https://learn.uni.lu/