Convergence of Skill Profiles for Cognitive Diagnosis Models and other Multidimensional Scaling Approaches: An Empirical Illustration with a Diagnostic Mathematics Assessment

Olga Kunina & Oliver Wilhelm

Institute for Educational Progress (IQB)

Humboldt University Berlin

André A. Rupp

Department of Measurement, Statistics, and Evaluation (EDMS)

University of Maryland

This research was supported by Grant # RU 424/3-1 from the German Research Foundation (DFG)

Presented at the annual meeting of the Psychometric Society (IMPS) in Durham, NH, June 29 – July 2, 2008

Research Motivation

Diagnostic classification models (DCMs) have been proposed in recent years as powerful statistical tools for providing fine grained feedback about multiple skills of learners.

The success of applying these models rests on having appropriately designed diagnostic assessments at hand that contain a sufficient number of items to index each skill.

Research Motivation

However, few empirical examples exist that illustrate how DCMs can provide an added value over alternative multidimensional calibrations of assessment data in these contexts.

Moreover, studies that employ DCMs rarely seek to cross-validate the resulting skill profiles with secondary variables.

Hence, practitioners have few guidelines and case studies for learning about the benefits and limitations of DCMs viz-à-viz scaling alternatives.

Research Questions

1. Can pilot data from a newly developed diagnostic assessment for arithmetic ability in elementary school be successfully scaled with DCMs?

2. How do the model fit and the resulting parameter interpretations differ across the best-fitting DCM and the best-fitting traditional multidimensional scaling model?

3. What lessons can be learned from these analyses about the suitability of the design of the diagnostic assessment and the benefits and limitations of DCMs and alternative scaling models?

Diagnostic Classification Models

Properties of DCMs

Diagnostic classification models…

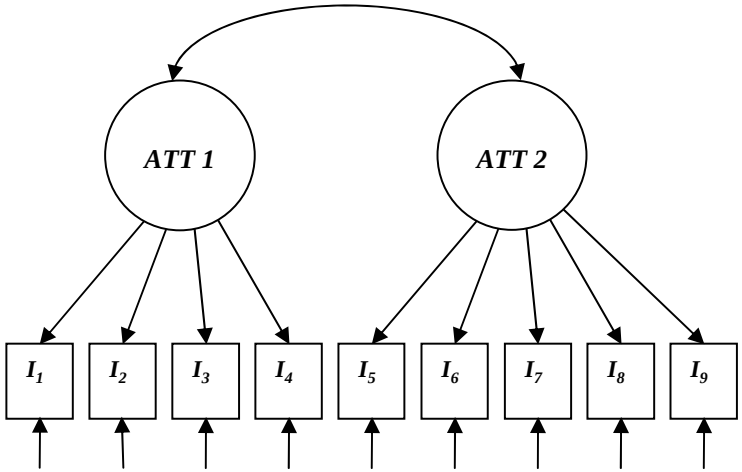
- are restricted latent class models
- are multidimensional latent variable models with discrete latent variables
- provide classifications of learners according to multiple skills
- are designed to handle within-item multidimensionality efficiently
- include compensatory and non-compensatory interactions of latent variables

Properties of DCMs

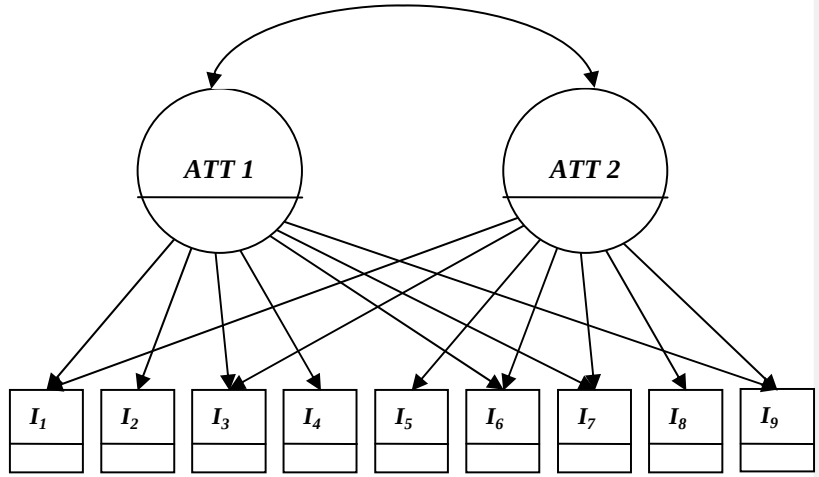
Diagnostic classification models…

- can handle dichotomous and polytomous response variables
- can be estimated with frequentist and Bayesian approaches
- can handle observed and latent mixtures
- can handle nested data structures
- are currently being unified under log-linear estimation frameworks

Illustration of DCMs



Two-dimensional CFA Model with Simple Loading Structure
(Exploratory or confirmatory)



Two-dimensional DCM with Complex Loading Structure
(Confirmatory)

Test & Data-collection Design

Data-collection Design

Sample

n = 241 children in grade 3

n = 223 children in grade 4

All children were selected from the same k = 6 schools located in the federal states of Germany that were part of a network of schools participating in several large-scale assessment studies

Test Administration

Standardized instructions

Trained test administrators (trained graduate students)

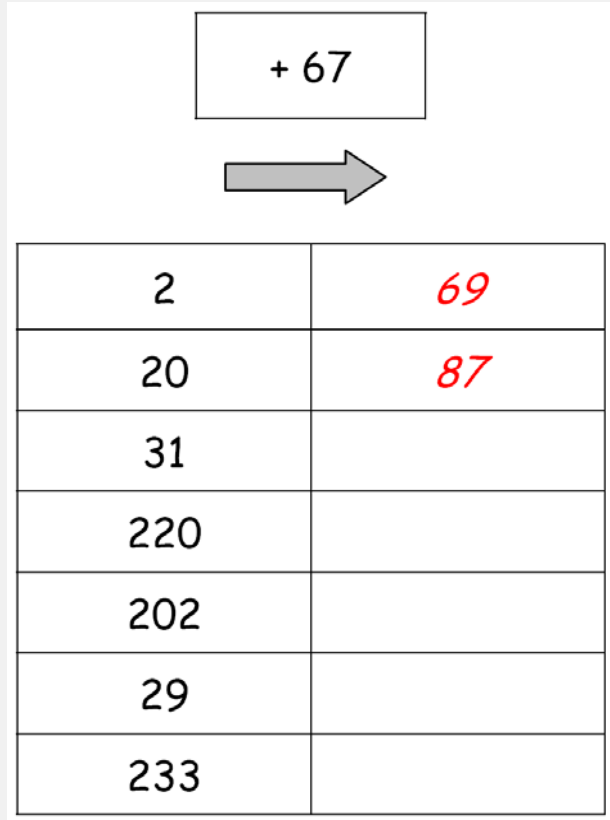
Total testing time = 90 mins. (2 school periods) with a 10-minute break

Test Design (Part I)

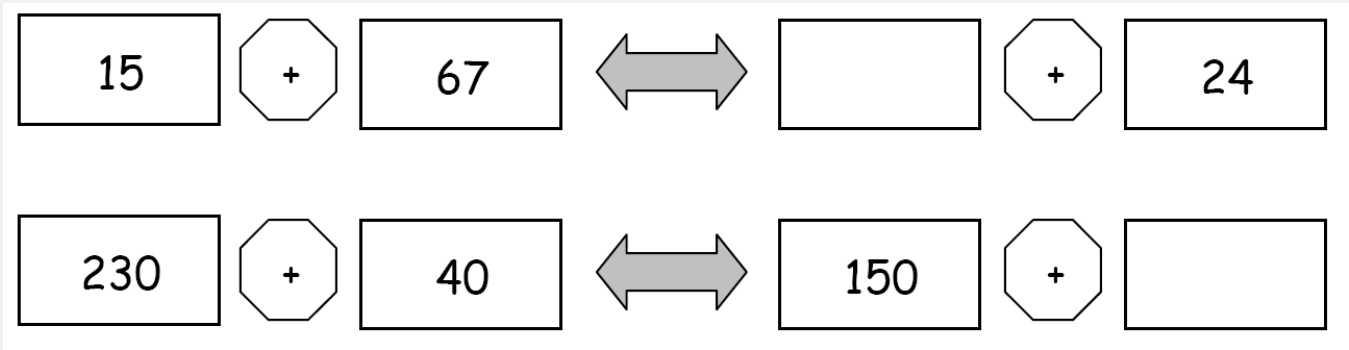
	# Items
A) Basic Arithmetic Skills	
1. Addition	14
2. Subtraction	14
3. Multiplication	12
4. Division	12
5. Combined Skills	20
B) Basic Processing Skills	
1. Knowledge of the decimal system	14
2. Ability to perform carry-over operations	20
3. Ability to perform inverse operations	50
4. Knowledge of the number space over 100	29
C) Modelling Skills (for Contextualized Problems)	39

Sample Items (Context-free)

7	+	5	=	*12*
12	+	12	=	24
20	+	8	=	
2	+		=	101
23	+		=	76
	+	84	=	190



2	*69*
20	*87*
31	
220	
202	
29	
233	



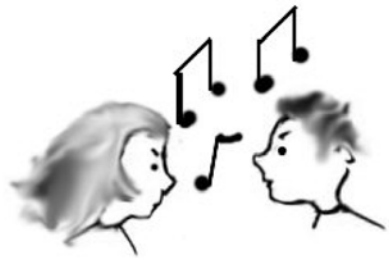
Sample Item (Contextualized)

T_AS5

Marlen has 55 songs on her MP3-Player. Leon has 15 songs less than Marlen. When they go through their song lists, they realize that the lists have 10 songs in common.

How many different songs do Marlen and Leon have altogether?

Marlen and Leon have _____ different songs altogether.



Test Design (Part II)

Context-free / Basic Arithmetic Items

	Total	Carry-over	Inverse	Number Space
Addition	22	10	16	10
Subtraction	22	10	14	7
Multiplication	20	0	18	6
Division	20	0	18	7

Statistical Modelling Approach

1. Model on global arithmetic ability (context-free items; IRT)

2. Model on global arithmetic ability (contextualized items; IRT)

3. Model on basic arithmetic skills (context-free items; IRT & DCM)
   (addition, subtraction, multiplication, division)

4. Model on basic arithmetic skills (contextualized items; IRT & DCM)
   (addition, subtraction, multiplication, division)

5. Higher-order model (context-free items; IRT, & DCM)
   (arithmetic ability + addition, subtraction, multiplication, division)

6. Higher-order model (contextualized items; IRT, & DCM)
   (modelling ability + addition, subtraction, multiplication, division)

7. Model on processing skills (context-free items; IRT & DCM)
   (addition, subtraction, carry-over, inverse, number space)

Compare calibrations across grades to separate calibrations for grades 3 and 4.

DCM Results
(General Diagnostic Rasch-type Model)

DCM Results
(General Diagnostic Rasch-type Model)

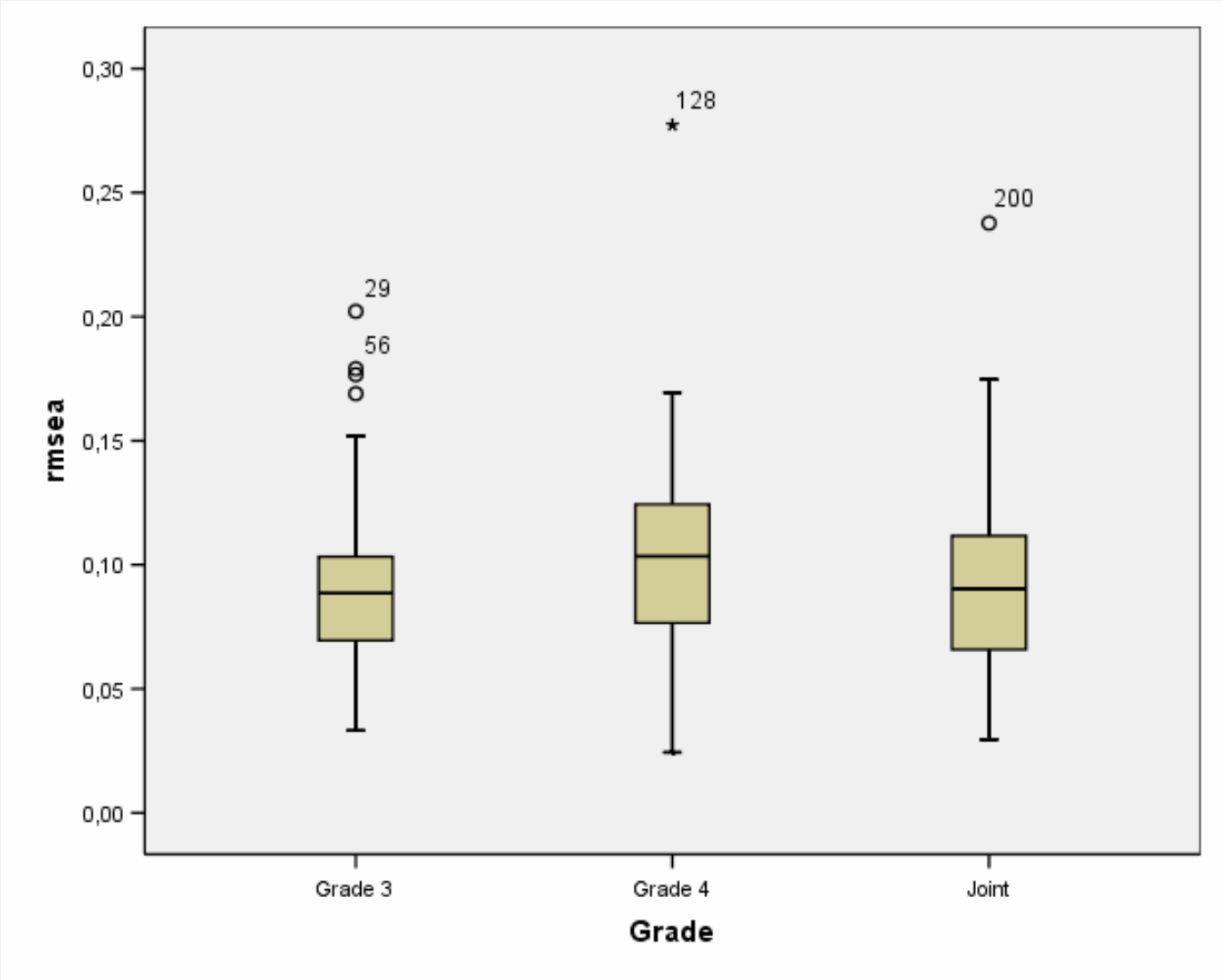
	Addition	Subtraction	Multiplication	Division
Addition	.89 / .90	.65	.68	.58
Subtraction	.73	.90 / .92	.65	.62
Multiplication	.23	.35	.91 / .96	.78
Division	.34	.23	.70	.89 / .94

Reliabilities = diagonal, Grade 3 = lower off-diagonal, Grade 4 = upper off-diagonal

Skill Information

Grade	Addition	Subtraction	Multiplication	Division	% Children
3	1	1	1	1	25.8
	0	0	0	0	20.1
	1	1	0	0	23.2
	1	1	1	0	11.9
	1	0	0	0	7.3
	72.2	62.5	47.6	33.5	% Mastery
4	1	1	1	1	37.5
	0	0	0	0	30.6
	1	0	0	0	8.0
	1	1	0	0	4.6
	1	0	1	1	4.3
	1	1	1	0	3.5
	59.5	51.6	48.2	48.5	% Mastery
Joint	0	0	0	0	31.0
	1	1	1	1	26.3
	1	1	0	0	11.7
	1	0	0	0	9.6
	1	1	1	0	6.1
	59.7	49.0	42.2	35.1	% Mastery

Fit Information



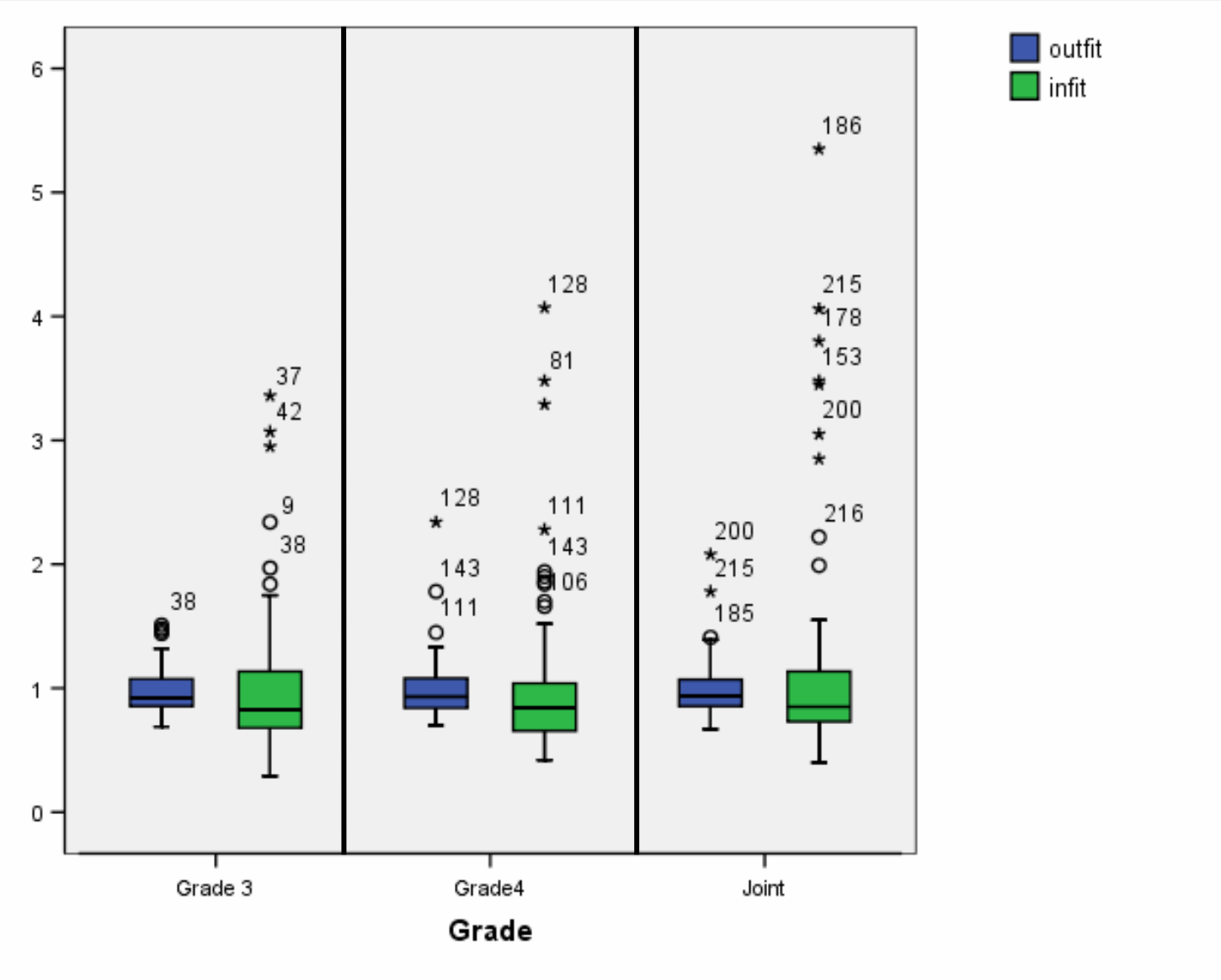
IRT Results
Multidimensional Rasch Model

IRT Results
Multidimensional Rasch Model

	Addition	Subtraction	Multiplication	Division
Addition	.89 / .91	.95	.75	.73
Subtraction	.94	.88 / .99	.75	.74
Multiplication	.51	.52	.90 / .99	.99
Division	.39	.40	.95	.92 / .99

Reliabilities = diagonal, Grade 3 = lower off-diagonal, Grade 4 = upper off-diagonal

Fit Information



Consistency of Results between DCM & Multidimensional Rasch Model

Skill	Mean Correlation $(\hat{\theta}_{PV_k}, \hat{\alpha})$ Grade 3	Mean Correlation $(\hat{\theta}_{PV_k}, \hat{\alpha})$ Grade 4	Mean Correlation $(\hat{\theta}_{PV_k}, \hat{\alpha})$ Joint
Addition	.74	.77	.75
Subtraction	.75	.77	.76
Multiplication	.76	.78	.76
Division	.72	.79	.75

Correlations with School Grades in Mathematics for Grade 3 (N = 216) and Grade 4 (N = 188)

	Grade 3		Grade 4	
	DCM	IRT	DCM	IRT
Addition	-0.36	-0.47	-0.38	-0.47
Subtraction	-0.39	-0.48	-0.41	-0.49
Multiplication	-0.37	-0.39	-0.32	-0.44
Division	-0.33	-0.34	-0.36	-0.44

Conclusions

- *Careful test design* paid off reasonably well for these analyses
- *Multidimensional scaling* at least moderately successful with IRT models and DCMs
- *Moderate consistency* between IRT estimates and DCM estimates
- *Empirical separation* of addition / subtraction and multiplication / division challenging with current items and models
- *Additional fine-tuning* of analyses will likely lead to lower-level scaling for diagnostic feedback about sets of arithmetic skills

Outlook

- *Replication* with a larger sample size from field trial
- *Finer look* at local model misfit with subsequent model refinement
- *Additional comparisons* of higher-order IRT models, SEMs, and DCMs
- *Additional comparisons* of multi-level IRT models and DCMs
- *Cross-validation* of from the person parameters in IRT models, SEMs, and skill profiles in DCMs via secondary measures

Contact Information

Olga Kunina & Oliver Wilhelm

Institute for Educational Progress (IQB)

Humboldt University Berlin

E-mail: olga.kunina@iqb.hu-berlin.de

André A. Rupp

Department of Measurement, Statistics, and Evaluation (EDMS)

University of Maryland

E-mail: ruppandr@umd.edu

CFA Model Comparisons

Model	Skils / Description	n	# Items	χ^2	# Pars	CFI	TFI	RMSEA	SRMR	WRMR
1	A, S, M, D (*incl. items that combine skills*)	464	72	936.67	90	.93	.97	.09	.11	1.68
2	A, S, M, D (*reduced*)	464	43	565.67	56	.96	.98	.07	.10	1.48
3	Math & A, S, M, D (*reduced*)	464	43	765.94	47	.95	.97	.10	.13	1.91
4	A, S, M, D (*reduced*) & Contextualized Items (*one-skill)*	464	55	1559.40	61	.88	.93	.14	.19	2.60
5	A, S, M, D (*reduced*) & Decimals (*reduced*)	185	61	240.37	71	.96	.98	.09	.13	1.29
6	Math & A, S, M, D (*reduced*) & Decimals (*reduced*)	185	61	260.55	66	.95	.97	.10	.14	1.37
Et al.	Models with A, S, Carry-over, Inverse, Number Space	*No convergence*								

Test Design (Part III)

Contextualized / Problem-solving Items

	Total	Carry-over	Inverse	Number Space
Addition	23	4	5	0
Subtraction	23	6	5	0
Multiplication	16	2	2	0
Division	16	0	0	0
Modelling	13	0	0	0