

Exploring Construct Equivalence of SAQ and MCQ Items Using IRT

How Uniformly Do Different Item Types Measure Reading Comprehension among 6th Graders Learning French?

Peter Lenz, Malgorzata Barras, Katharina Karges / Task Lab Project

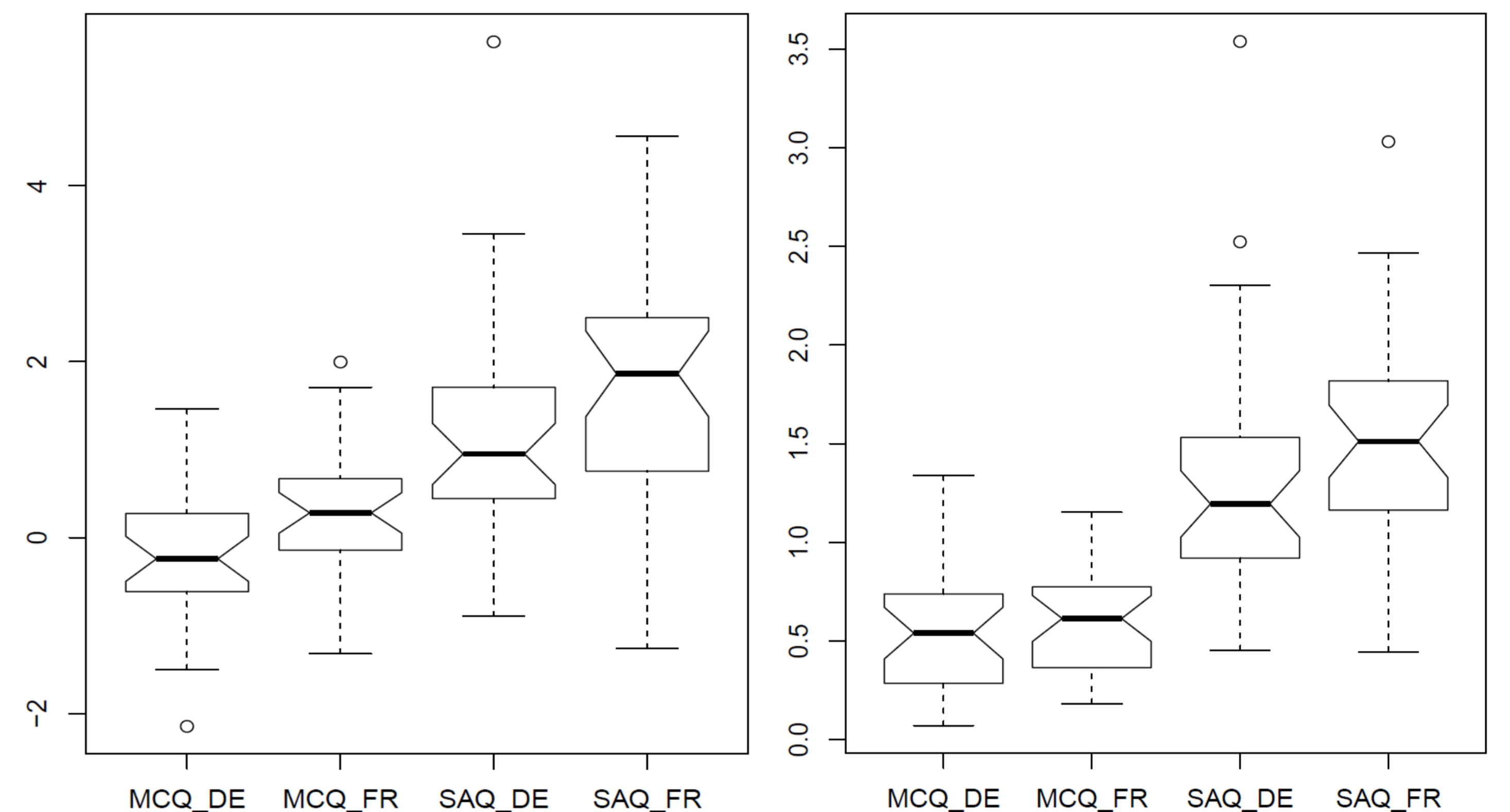
READING TASKS USED

'Same' task • 2 item types • 2 languages (for questions & answers)

N ≈ 600 6th graders in 35 classes solved a sample of tasks in all 4 variants.

HOW DO SAQ AND MCQ ITEMS 'BEHAVE'? FINDINGS FROM 1-DIMENSIONAL ANALYSIS

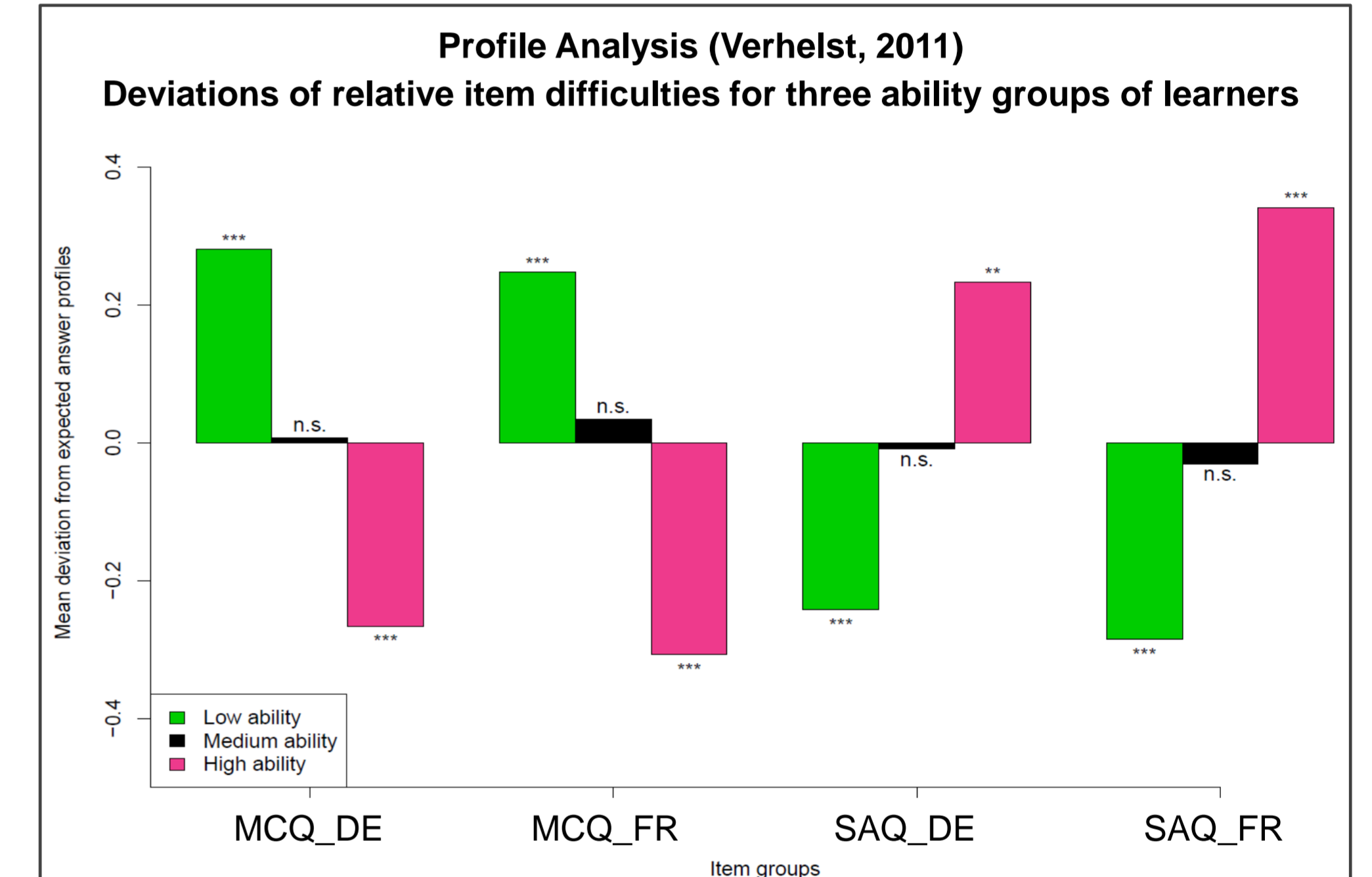
Item Difficulties on Item Type X Language Item Discriminations on Item Type X Language



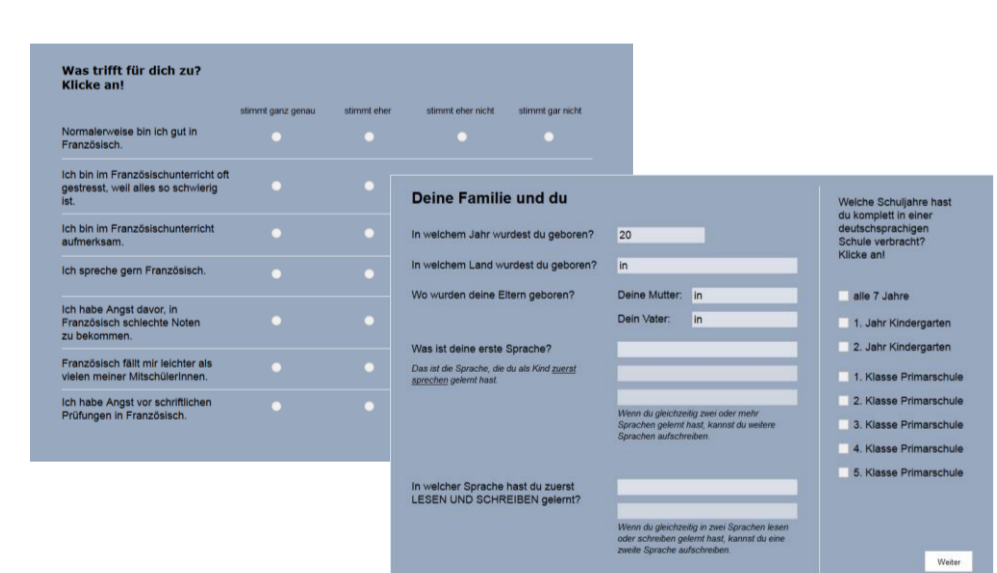
As expected, a 2PL model, accounting for individual item discriminations fits better:
Rasch Mod. vs. 2PL Mod. (dev. diff): $\chi^2(328, 125), p < 0.001$
2PL model with 2 slope groups (SAQ = 1.27; MCQ = 0.58):
• better deviance, AIC and BIC than Rasch
• worse deviance, better AIC, BIC than unrestricted 2PL model

Some Observations

- SAQ items are relatively more difficult (*no problem*).
- The SAQ items together form a longer measurement scale, i.e. they differentiate better between students.
- The individ. SAQ items have (much) higher discrim. => There are 2 item groups present (MCQ vs. SAQ)
- For the low ability group of test-takers, SAQ items are relatively more difficult than for the high ability group – the opposite is true for MCQ items. => Violates the principle of specific objectivity.



PREDICTORS USED



Student Questionnaire

- Gender
- (Rom, lang, background)
- Motivation (enjoyment)
- Motivation (ought)



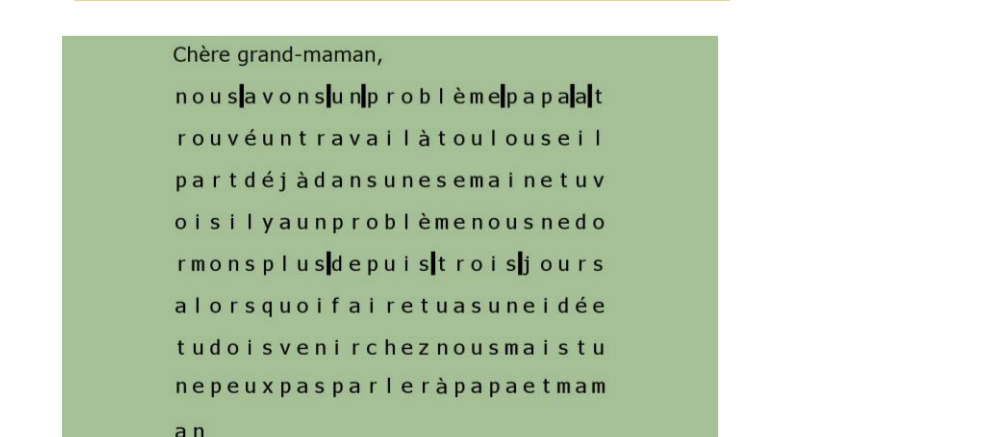
Backward Digit Span Task
Working memory/ processing



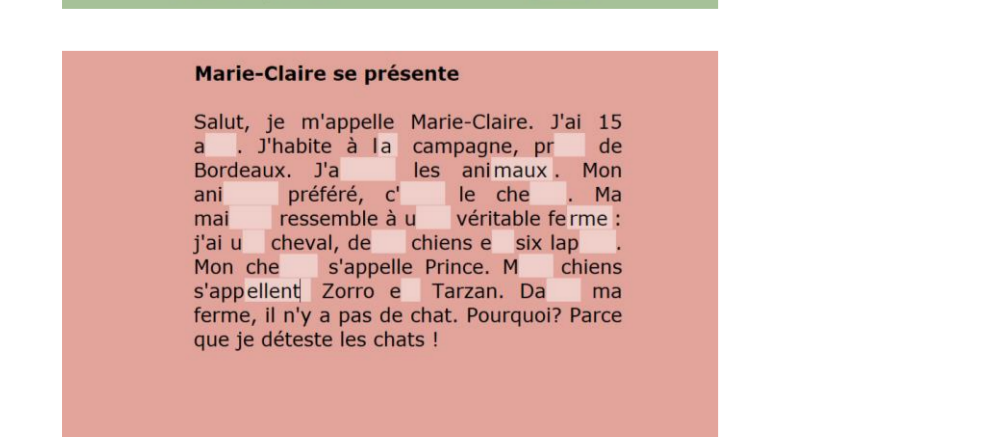
Sight-word recognition
Word decoding (gestalt)



Yes-No Task
Vocabulary breadth (receptive)



Text segmentation
Morpho-syntax & integrative measure



C-Test
Integrative measure / written text reconstruction

Predictor matrix was completed through imputation using the Amelia II R package (max. 10.7 % missings)

DOES SUCCESS ON MCQ AND SAQ HAVE THE SAME PREDICTORS? – LATENT REGRESSION ON 2 DIMENSIONS

A 2-dimensional (per item type) model fits better than the 1-dimensional model
Mod.2PL.1Dim vs. Mod.2PL.2Dim (dev. Diff.): $\chi^2(10.85, 1), p < 0.001$

Results of latent ('error-free') regression on 2 dimensions:

Predictors	Predictor type	Dim 1 SAQ	Dim 2 MCQ
Gender: male	dummy	0.095	-0.115
Romance lang. background	dummy	0.512	0.112
Motivation: enjoyment	z-std.	0.181	0.109
Motivation: ought	z-std.	0.038	-0.013
Backward digit span	z-std.	0.150	0.145
Sight-word recognition	z-std.	0.159	0.145
Yes-No Test (recognise word)	z-std.	0.142	0.292
Segmentation task	z-std.	0.406	0.293
C-Test	z-std.	0.221	0.156

Some Observations

- Known correlates of better language knowledge predict success on SAQ items particularly well: Integrative measures, Motivation (enjoyment ≈ intrinsic), and a romance family language background (13.6% of sample).
- The strictly receptive Yes-No word recognition test is a better predictor for success on MCQ. The possibility of success through guessing may be a commonality (despite a correction for guessing made on YNT).

Discussion

- It seems desirable to be able to pinpoint more specific component knowledge & skills – which ones?
- What item characteristics should be taken into account for a rigorous person-item explanatory model?