A Test Assembly Model for MST

IACAT, 4th Oct. 2011

Angela Verschoor, Cito,

Ingrid Radtke, Vox,

Theo Eggen, Cito/University of Twente

angela.verschoor@cito.nl

now you know

Multistage Testing

Development of a MST:

- Developing items

- Pretesting

- Assembly of the MST:
    - Multiple modules
    - Overlap between modules?
    - Automated Test Assembly models can be useful

ATA models for MST

Derived from ATA for parallel tests:

- Each module is a subtest, parallel

Specifications for each module:
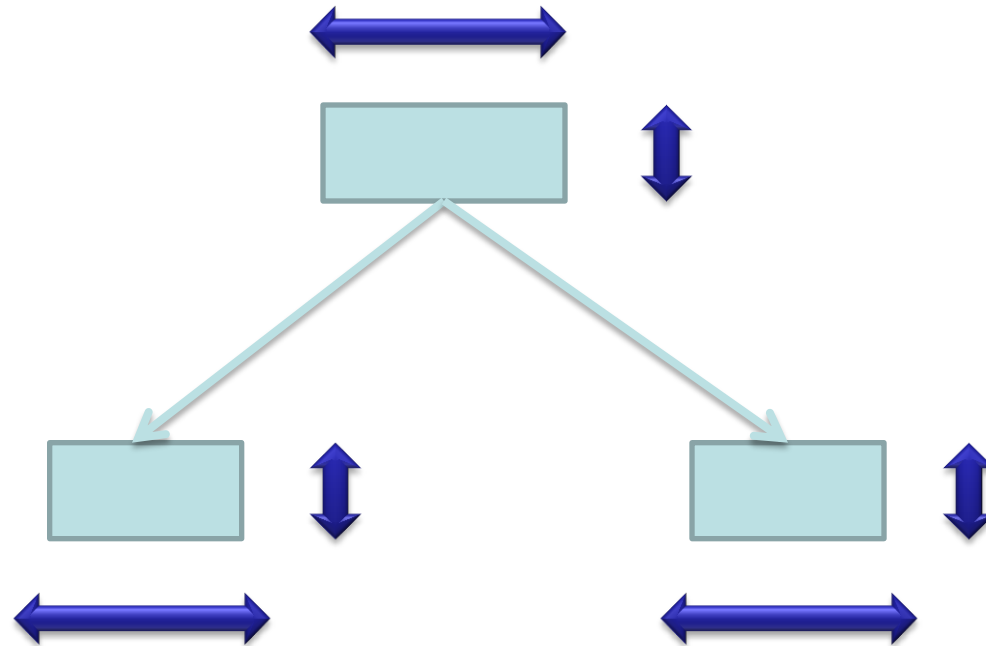
- TIF target
- Content, length, etc

Specifications are fixed, not part of the optimization model

ATA for MST

Specifications:

Two problems

- What is a suitable TIF target?

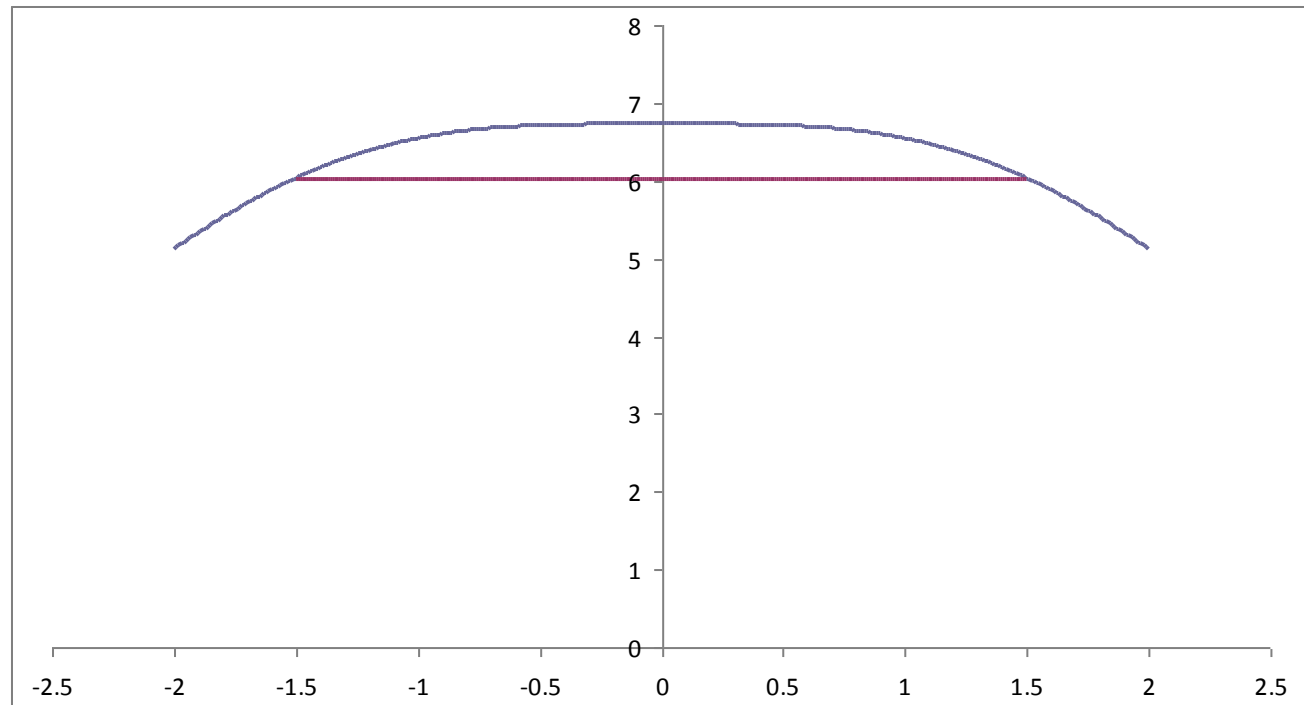- What is the optimal module length?

now you know

TIF target

- Purpose of the test!

- Assumption:

   ability estimation,

   minimization of measurement error.

- Optimize TIF target, or
- Optimize constraints for ATA model

TIF target (2)

Flat TIF target:

=> Focal points

TIF target (3)

Flat range -a ..a:

focal points -b and b:

$$b = 0 \qquad \text{for } a \leqslant \log(2 + \sqrt{3})$$

$$b = \log\left(\frac{1 - 6e^{2a} + e^{4a} + \sqrt{1 - 16e^{2a} + 30e^{4a} - 16e^{6a} + e^{8a}}}{2(e^a + e^{3a})}\right) \qquad \text{for } a > \log(2 + \sqrt{3})$$
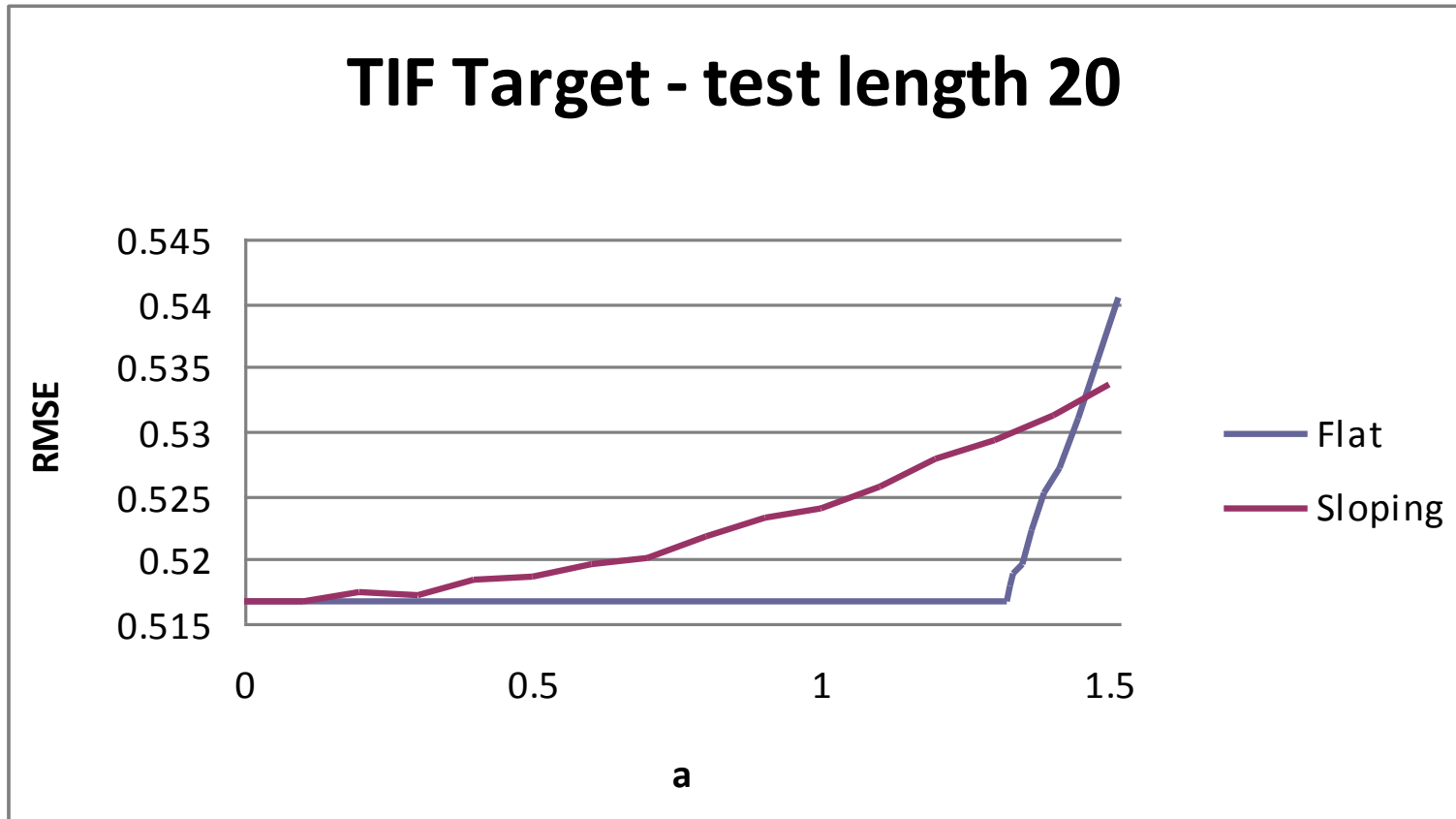
or: range of difficulties, gently sloping TIF

TIF Target - test length 20

Routing

Routing module: classification

- Minimize variance in subpopulations
  - (Roughly) equal size
  - Minimal classification error

- Routing decision based on $\hat{\vartheta}$
- Minimization of error => maximization of information at decision point

now you know

Routing: an example

Routing module: two Rasch items:

Three $\widehat{\vartheta}$ s: associated with score 0, 1, 2
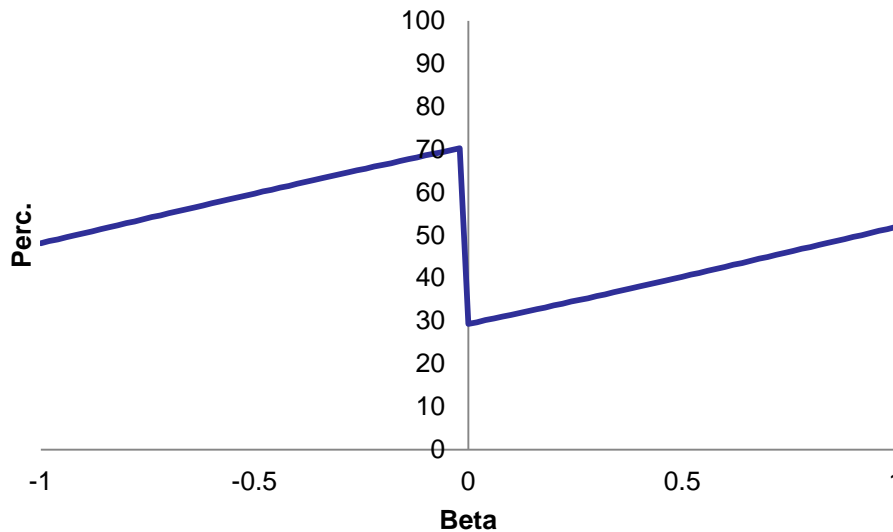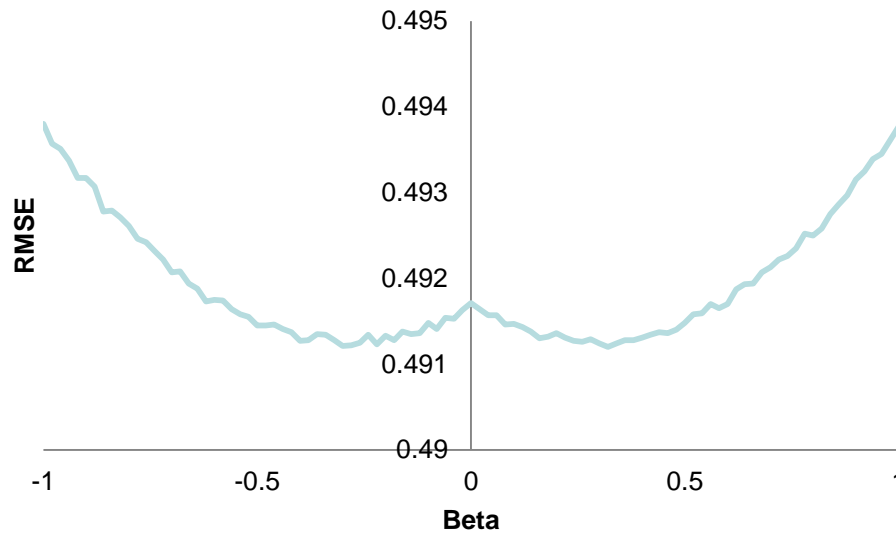
Two most informative items:

70% - 30%

50% - 50%:

b = -0.92 or b = 0.92

now you know

Routing: an example (2)

Routing Module Length

Minimization of classification error:

- extending routing module
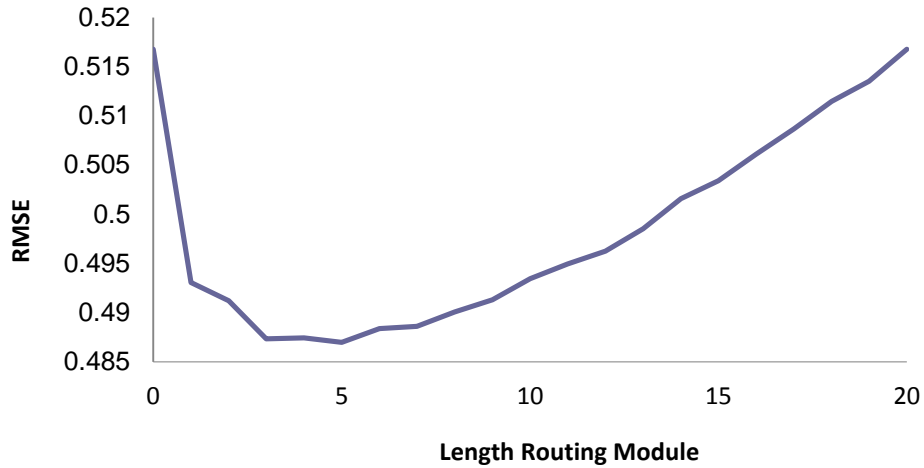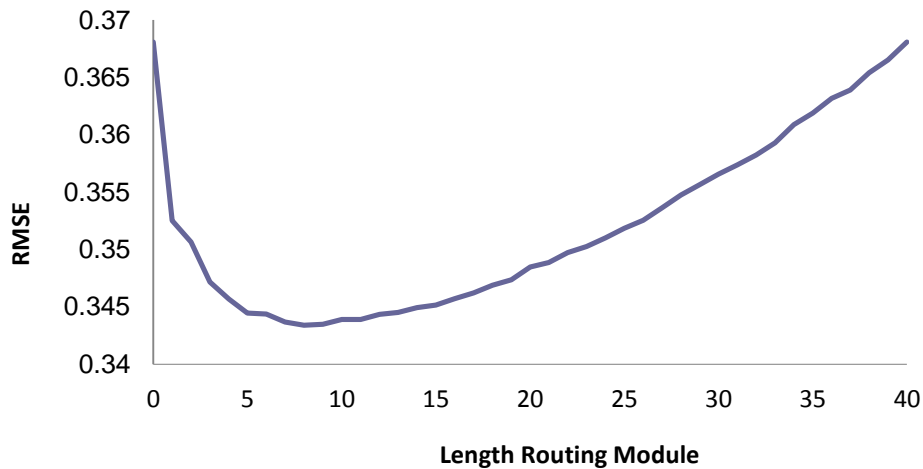
But: fewer items available for targeted modules

now you know

Routing Module Length (2)

Test length 20
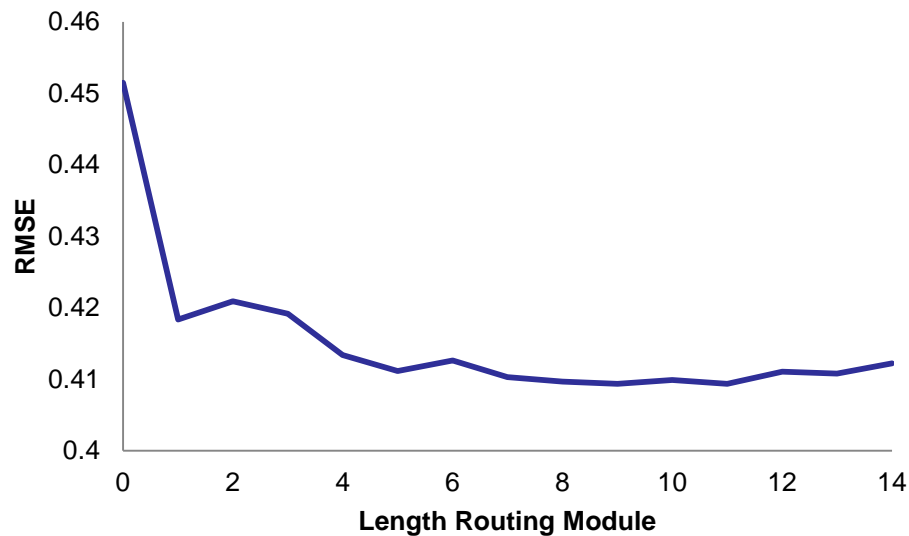

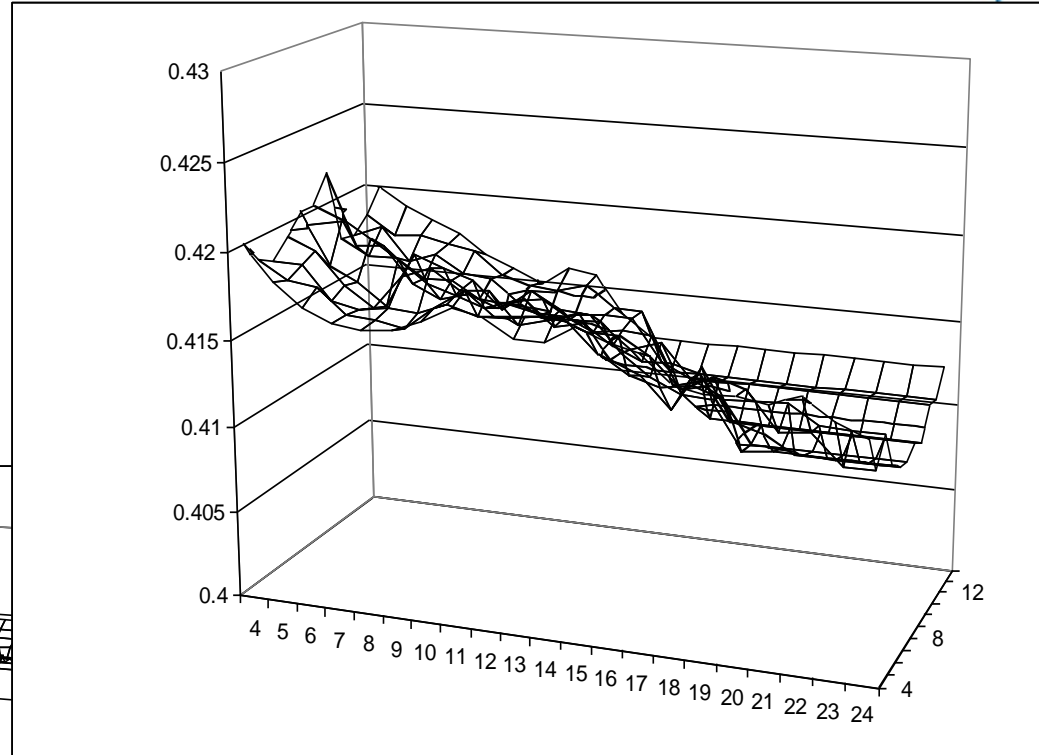**Test length 40**
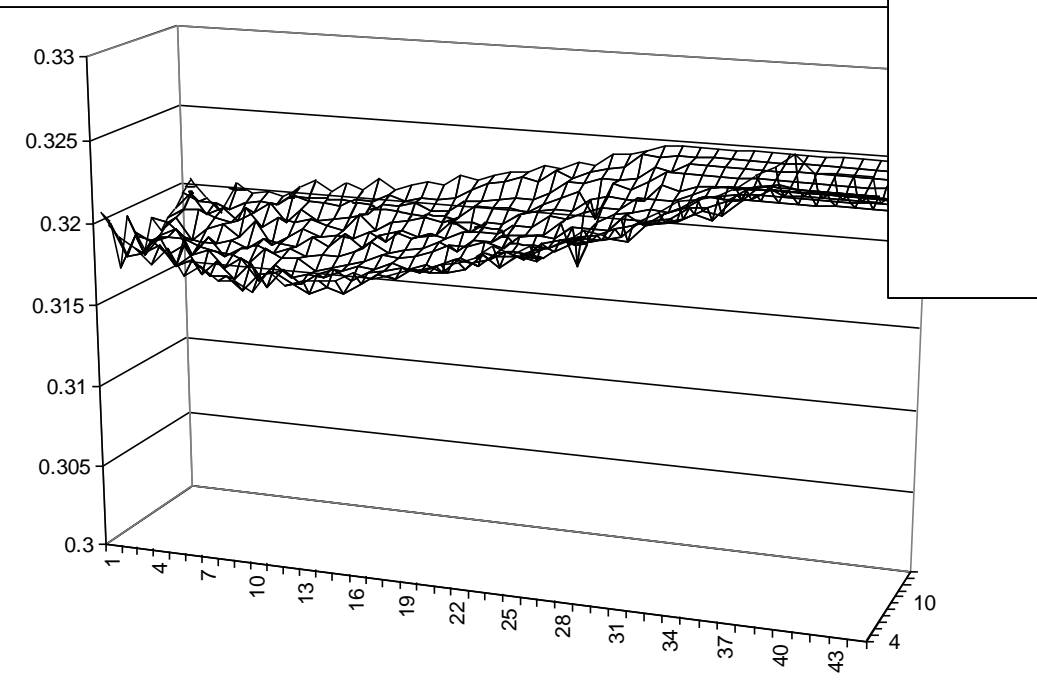
A real life example

Vox: adult education, Arithmetic Test

Item pool 99 items, 2PL

Desired test length: 28

Three stages

Conclusions

This study is just a short exploration in the matter of optimization of a MST

Extremely hard or maybe impossible to chart influence of item pool and test specifications on optimization process

Simulations are very helpful in finding an acceptable MST

cito now you know

Thank you

?

angela.verschoor @cito.nl