12th Annual UK Rasch User Group Meeting
March 23, 2018
Loughborough

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

# Investigating the effect of the threshold distribution in polytomous items on measurement error *in computer adaptive testing*

*Thomas Salzberger*[1]
thomas.salzberger@wu.ac.at

*Mike Horton*[2]
m.c.horton@leeds.ac.uk

[1] University of Economics and Business, Vienna, Austria (WU WIEN)

[2] University of Leeds, UK

# Background: Measurement Precision

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

$$SE(\beta) = \frac{1}{\sqrt{I(\beta)}}$$

Polytomous items provide more information (Linacre, 2005; Stone, 2008)

$$I_i(\beta) = \sum_{k=0}^{m} \left(k - E(\beta)\right)^2 P_k(\beta)$$

*(for equivalent formulae: Dodd and Koch, 1994; Muraki, 1993)*

**Computer adaptive testing**

Administer the smallest possible number of items while obtaining a predefined level of the standard error of measurement.
*Min # of items given S.E.*

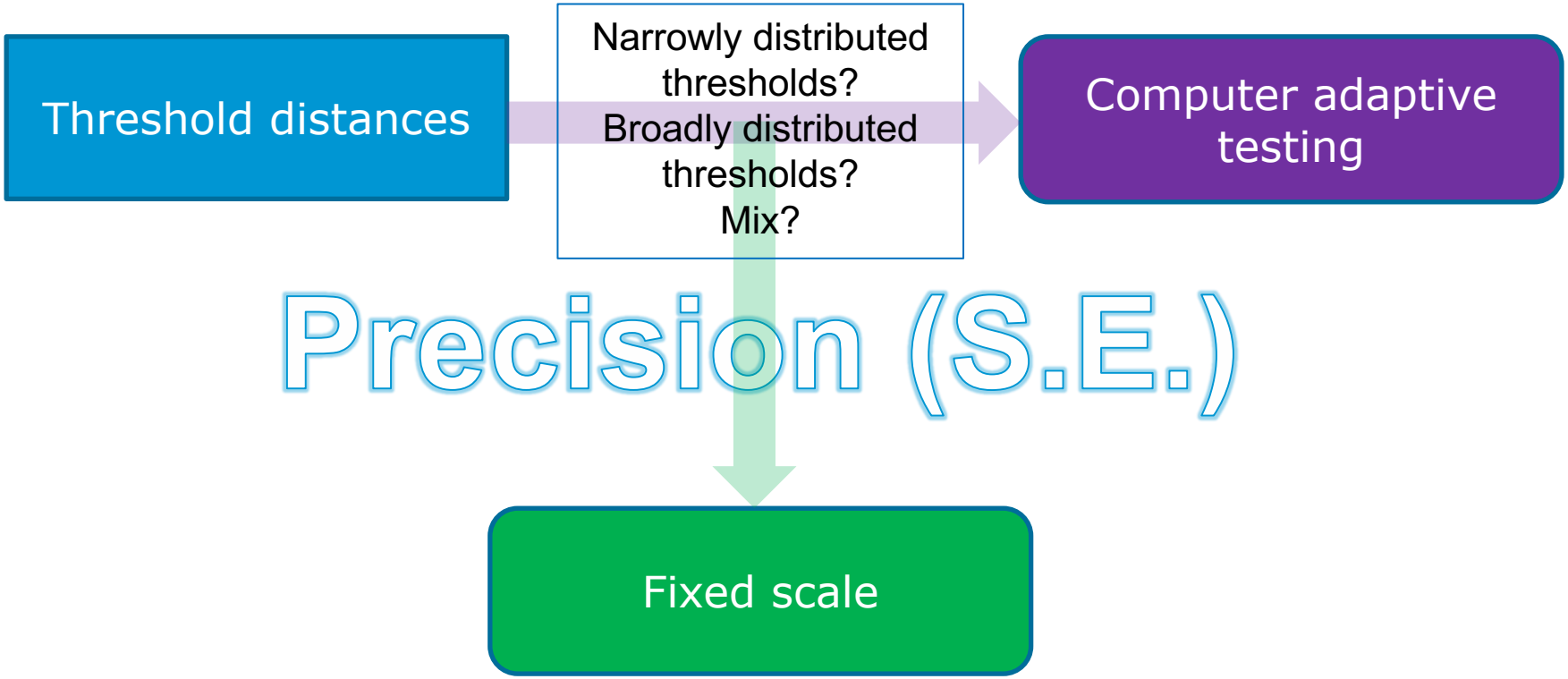
**Measurement precision**

**Threshold distances**

**Fixed scale**

Obtaining the smallest possible standard error of measurement while administering a fixed number of items in a scale
*Min S.E. given the # of items*

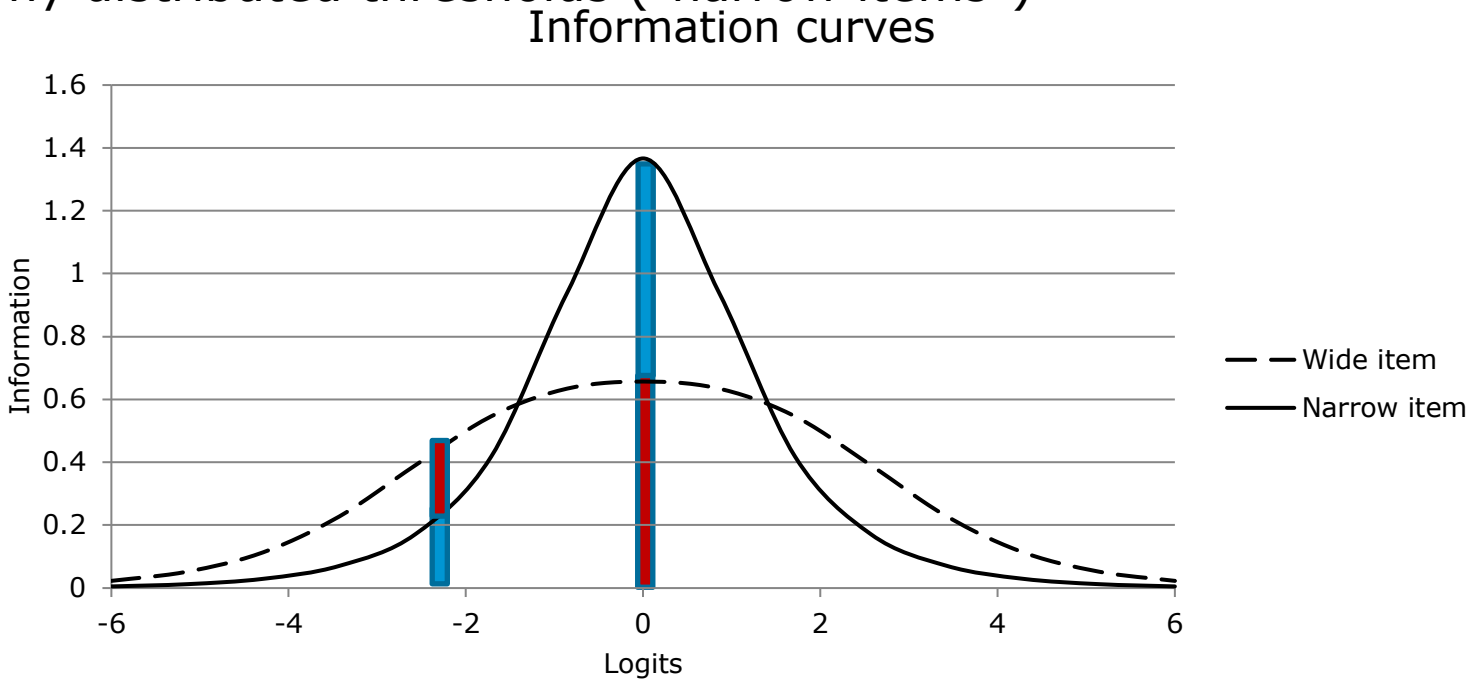
# Research Question



# Information Provided by Polytomous Items

Shape of information curve depends on threshold distribution

- Broadly distributed thresholds ("wide items")
- Narrowly distributed thresholds ("narrow items")

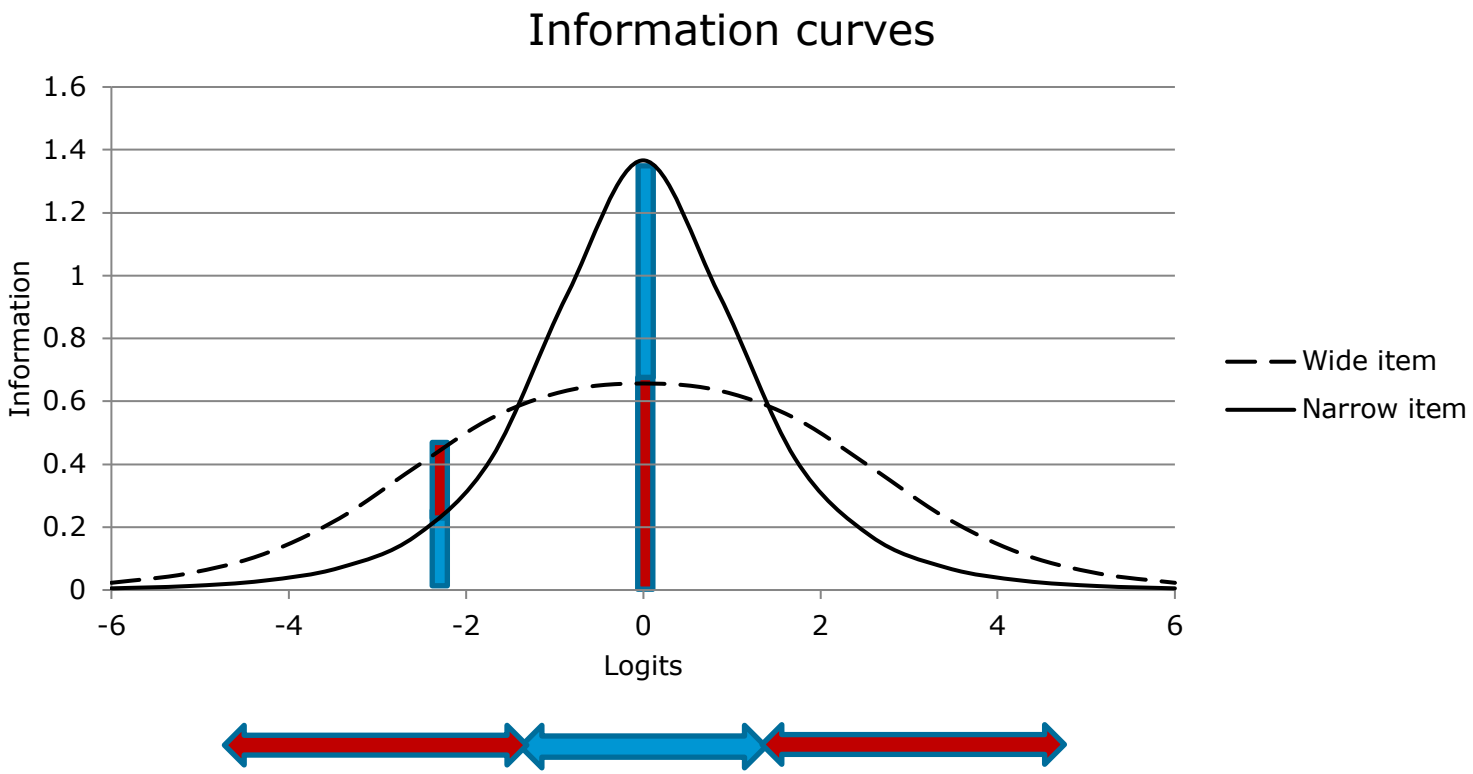



*Narrow item: thresholds at -0.75, -0.25, 0.25, and 0.75 (mean 0; Δ 0.5)*
*Wide item: thresholds at -2.25, -0.75, 0.75, and 2.25 (mean 0; Δ 1.5)*

# Research Proposition

- In CAT, at the beginning, when nothing is known about $\beta$, wide items should be more efficient.
  - Provide information over a broader range.
- Subsequently, when $\beta$ can be narrowed down, narrow items should be more efficient.
  - Provide more information at the item location.
- Caveat:
  - Information depends on person location and item location
  - Efficiency of wide versus narrow thresholds depends on person distribution
  - Targeting

# Information Provided by Polytomous Items



# Approach

Computer adaptive testing

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Item bank consisting of 16 items
  - Mean 0, evenly distributed between -3 and +3
- CAT simulation using Firestar (Choi, 2009)
  - Stopping criterion: SE=0.3 (never achieved in simulations reported here)
- S.E. after each item administered (up to all 16 items)

# Approach

Computer adaptive testing

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

**Threshold distance**

- Fixed item overall locations between -3 and +3
  - Wide items (threshold $\Delta$ = 1.5) [1]
  - Narrow items (threshold $\Delta$ = 0.5) [2]

[1] implies a broader coverage of the latent continuum ($\pm$ 4.5) than [2] ($\pm$ 3.5)

  - Narrow items II ($\Delta$ = 0.5) with items located between -4.5 and +4.5

**# of response categories**

- Four versus five response categories

**Targeting**

- Targeted population (mean 0, sd 2) versus Mis-targeted population (mean 3, sd 2)



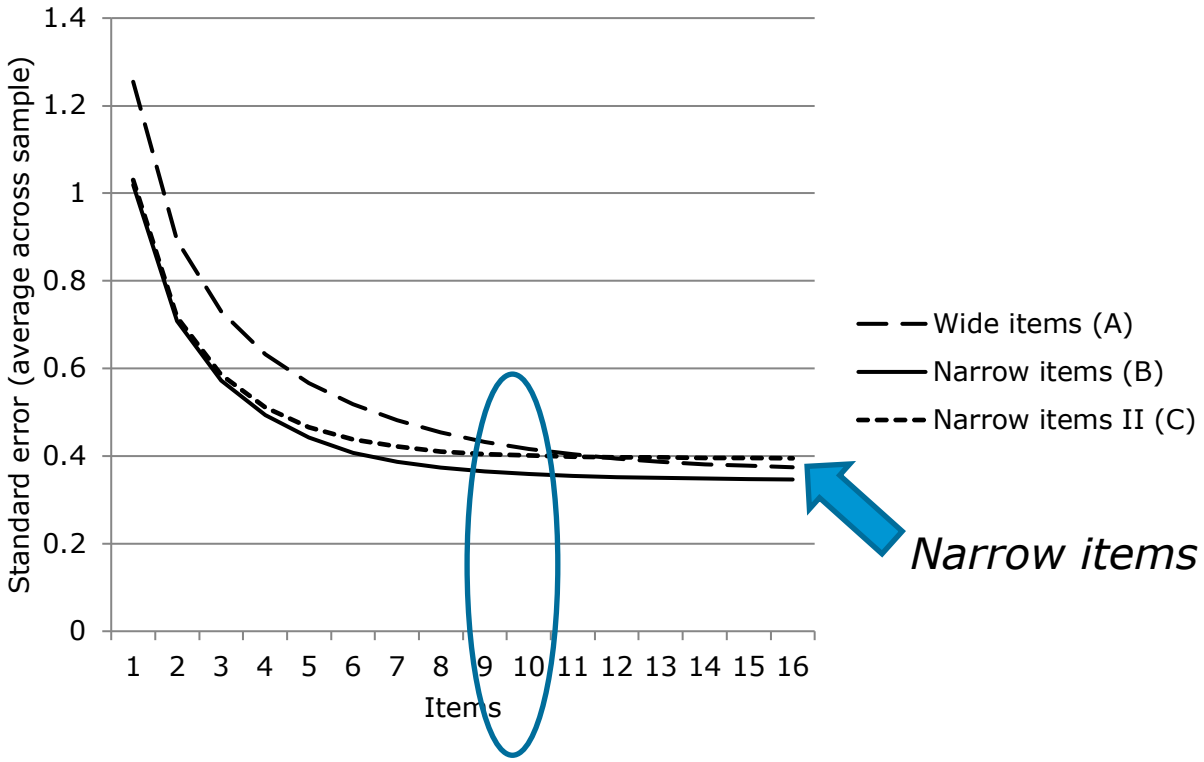
EFMD EQUIS ACCREDITED

# Settings

| Setting | Response categories (thresholds) | Item locations and threshold distance (16 items) | Respondent distribution |
|---|---|---|---|
| A – wide items | 5 (4) | -3 to +3; 1.5 | normal (0,2) |
| B – narrow items | 5 (4) | -3 to +3; 0.5 | normal (0,2) |
| C – narrow items II | 5 (4) | - 4.5 to + 4.5; 0.5 | normal (0,2) |
| D – wide items | 4 (3) | -3 to +3; 1.5 | normal (0,2) |
| E – narrow items | 4 (3) | -3 to +3; 0.5 | normal (0,2) |
| F – narrow items II | 4 (3) | - 4.5 to + 4.5; 0.5 | normal (0,2) |
| G – wide items | 5 (4) | -3 to +3; 1.5 | normal (3,2) |
| H – narrow items | 5 (4) | -3 to +3; 0.5 | normal (3,2) |
| I – narrow items II | 5 (4) | - 4 to + 4; 0.5 | normal (3,2) |
| J – wide items | 4 (3) | -3 to +3; 1.5 | normal (3,2) |
| K – narrow items | 4 (3) | -3 to +3; 0.5 | normal (3,2) |
| L – narrow items II | 4 (3) | - 4 to + 4; 0.5 | normal (3,2) |

# Results: 5 Response Categories, Targeted Population

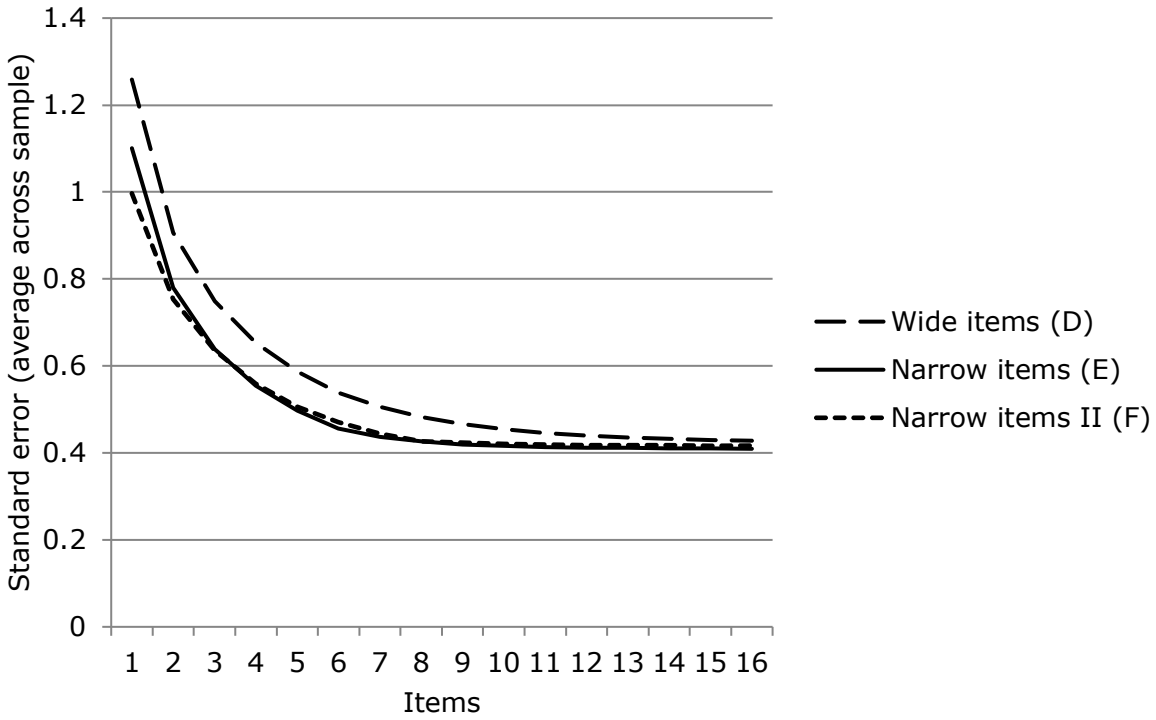
- Narrow items always perform better



# Results:
# 4 Response Categories, Targeted Population

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- Narrow items perform better

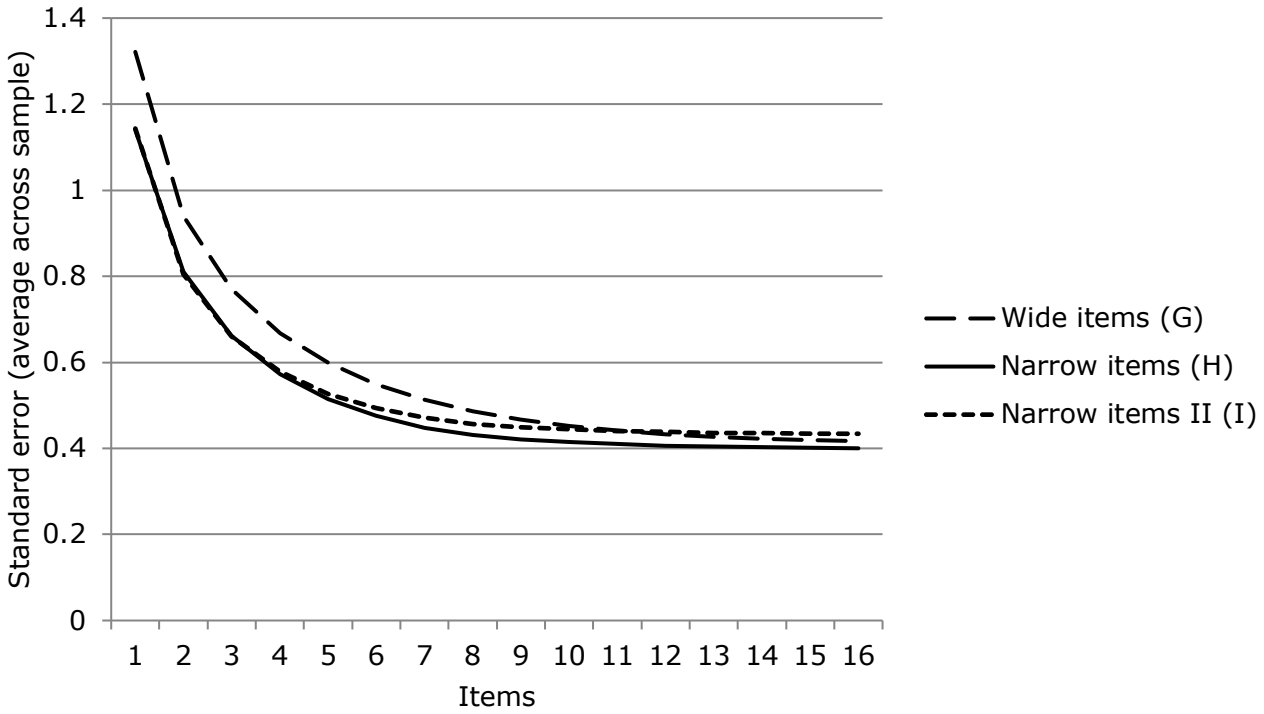





EFMD EQUIS ACCREDITED

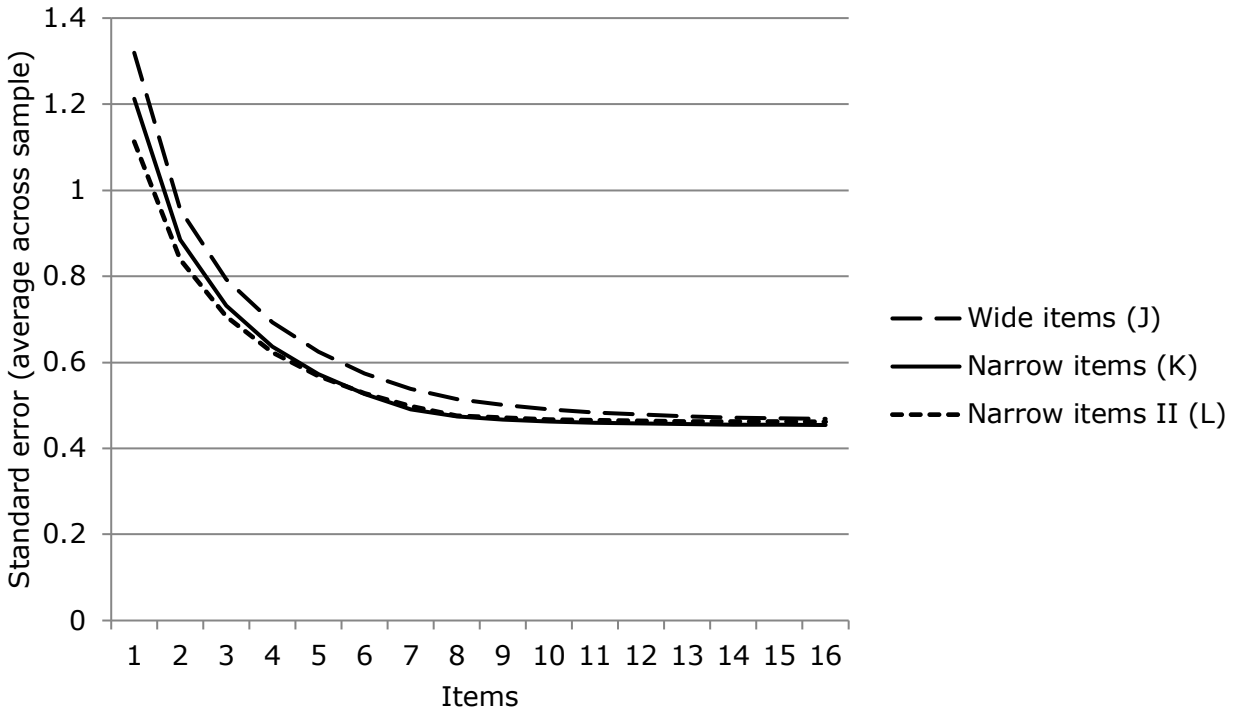
# Results:
# 5 Response Categories, Mis-targeted Population

- Narrow items perform better



# Results:
# 4 Response Categories, Mis-targeted Population

- Narrow items perform better



# Conclusions

- Narrowly distributed thresholds appear to contribute more to test information than broadly distributed thresholds
  - No indication that items with broadly distributed thresholds add to the efficiency of CAT
- Implication for fixed scales
  - Up to about 8 to 10-item-scales, narrowly distributed thresholds preferable
  - Beyond 10 items, there is hardly any difference
- In the scenarios investigated, effects do not depend on targeting

# Limitations and Further Research

- Conclusions based on average S.E.
  - Distribution?

- Realisation of optimum threshold distribution difficult in practice
  - Aiming for narrow thresholds might lead to indistinguishable response categories (implying threshold reversals)

# Limitations and Further Research

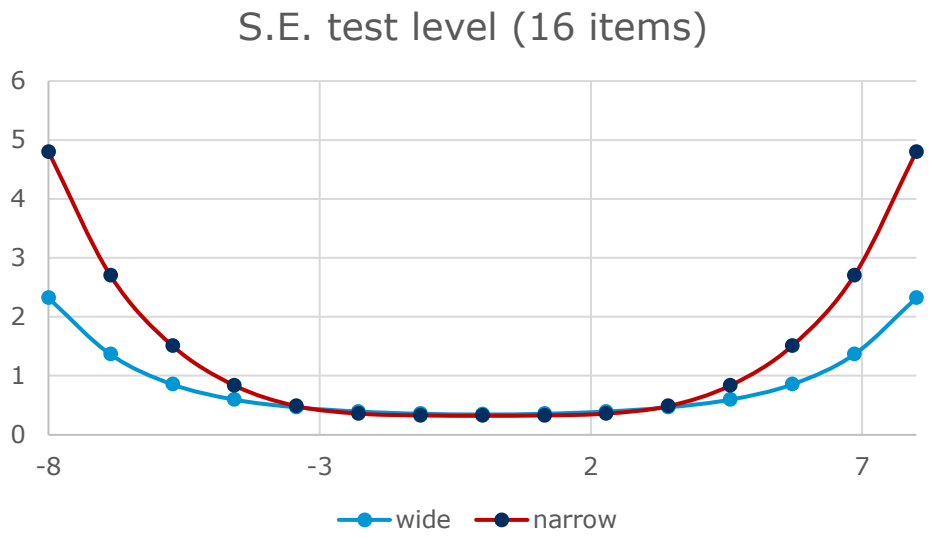
- Scenarios limited to 1.5 versus 0.5 distance between thresholds
  - Distance < 0.5?
- Scenarios limited to scales showing broad variation in terms of item locations
  - In practice, some scales exhibit relatively small variation
  - Are narrowly distributed thresholds still preferable?
- Scenarios based on a uniform distribution of item locations
  - In practice rarely the case

# Further Investigations Preliminary Results

- Very narrowly distributed thresholds (distance 0.25)
  - No noticeable difference between distance of 0.5 and 0.25
- Small variation in item locations (-1 to +1 [average distance 0.13] as opposed to -3 to +3 [0.40])
  - Broadly distributed thresholds perform better
  - Due to targeting and coverage
- Superiority of polytomous items with narrowly distributed thresholds contingent on sufficient coverage and good targeting
  - Broadly distributed threshold compensate for limited variation in item locations (from a precision point of view)
- Large variation in item locations and rather narrowly distributed thresholds appear to be optimum

# Further Research

- Computation of information and S.E. at the test level




| | Test (16 items) | |
|---|---|---|
| | S.E. | |
| beta | wide | narrow |
| -8 | 2.32021806 | 4.80288789 |
| -6.85714286 | 1.36455434 | 2.70233114 |
| -5.71428571 | 0.85538021 | 1.51071579 |
| -4.57142857 | 0.5970876 | 0.83720528 |
| -3.42857143 | 0.4665712 | 0.48714009 |
| -2.28571429 | 0.39564693 | 0.36017802 |
| -1.14285714 | 0.35732225 | 0.32918361 |
| -1.7764E-15 | 0.3449644 | 0.32335335 |
| 1.14285714 | 0.35732225 | 0.32918361 |
| 2.28571429 | 0.39564693 | 0.36017802 |
| 3.42857143 | 0.4665712 | 0.48714009 |
| 4.57142857 | 0.5970876 | 0.83720528 |
| 5.71428571 | 0.85538021 | 1.51071579 |
| 6.85714286 | 1.36455434 | 2.70233114 |
| 8 | 2.32021783 | 4.80288565 |

# Further Consideration

- Polytomous response scale adds information
  - Intensify item content (e.g., somewhat or strongly agree instead of just agree)
- But they do not contribute to characterising the latent variable
  - Item stem lends meaning to the latent variable