Scale construction and evaluation in practice: Factor analysis versus item response theory

Janke C. ten Holt

Marijtje A. J. van Duijn Anne Boomsma

University of Groningen

FCAP conference / June 26, 2008

university of groningen

Outline

university of groningen

Outline

university of groningen

FA and IRT

Statistical models in scale construction and evaluation:

- Factor analysis (FA)
- Item response theory (IRT)

university of groningen

FA vs IRT

Both latent variable (LV) models linking items to LVs (factors)

(Standard) FA

- Continuous item variables
- Linear relation between LV and items
- Model examples: CCFA, ECFA, OMG, PCA

IRT

- Categorical item variables
- Nonlinear relation between LV and items
- Model examples: Rasch, 2PLM, GRM, Mokken

Past research comparing FA and IRT

- Mathematically: Mehta & Taylor, 2006; Takane & De Leeuw, 1987; see also Kamata & Bauer, 2008
- Simulated data: Knol & Berger, 1991; Wirth & Edwards, 2007
- Empirical data: Glöckner-Rist & Hoijtink, 2003; Moustaki, Jöreskog, & Mavridis, 2004
- Simulated and empirical data: Jöreskog & Moustaki, 2001
- With regard to measurement equivalence: Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002; Reise, Widaman & Pugh, 1993

Central question

Research question:

What is done in practice and why?

university of groningen

Method

Review of 41 studies

- Concerning scale construction/evaluation
- Published in 2005 in
  - Psychological Assessment ($n = 13$)
  - European Journal of Psychological Assessment ($n = 13$)
  - Educational and Psychological Measurement ($n = 15$)

Method

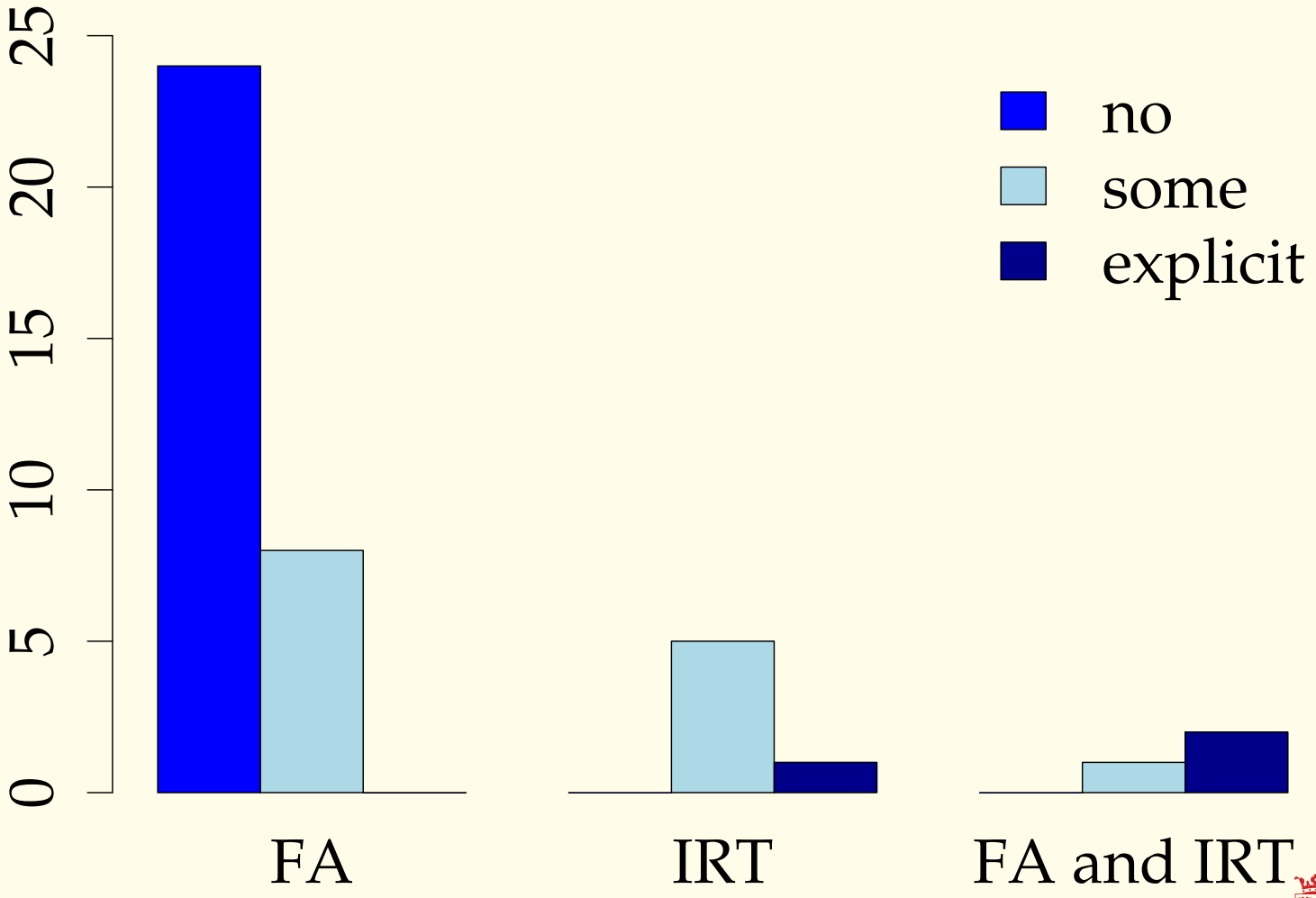
Review of 41 studies

- Concerning scale construction/evaluation
- Published in 2005 in
- Psychological Assessment ($n = 13$)
- European Journal of Psychological Assessment ($n = 13$)
- Educational and Psychological Measurement ($n = 15$)

Outline

university of groningen

Frequencies of motives in FA and IRT studies



university of groningen

Motives mentioning both FA and IRT

- Skewed item distribution -> Rasch models
- IRT better suited for dichotomous data

Outline

university of groningen

Number of categories

Table: No. of categories in studies applying FA, IRT, or both

		Type of applied analysis		
		FA ($n = 32$)	IRT ($n = 6$)	FA & IRT ($n = 3$)
No. of categories	2	4	1	1
	> 2	28	5	2

IRT not more often used for dichotomous data, as might have been suspected

Number of dimensions

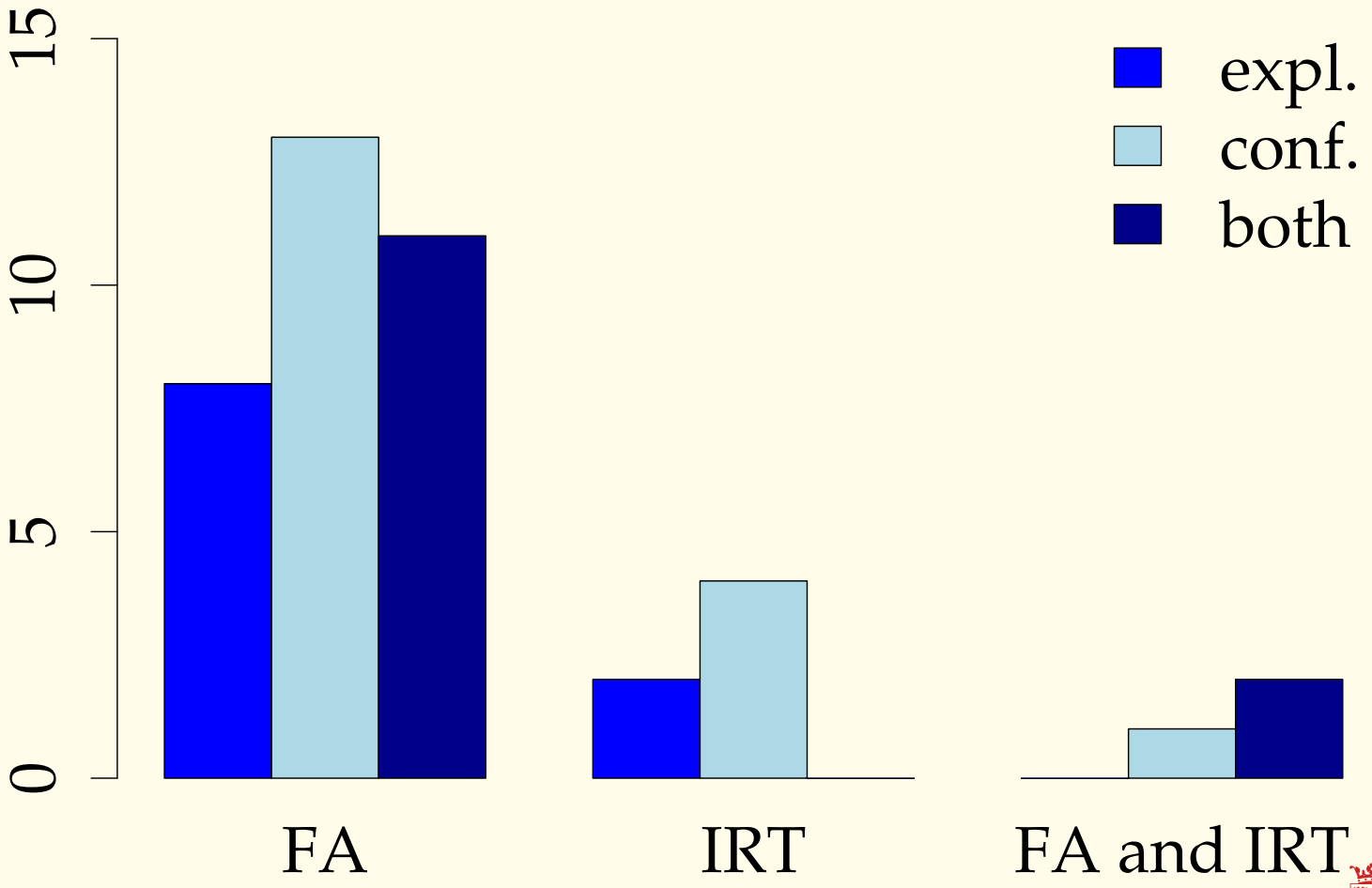
Table: No. of dimensions in studies applying FA, IRT, or both

		Type of applied analysis		
		FA ($n = 32$)	IRT ($n = 6$)	FA & IRT ($n = 3$)
No. of dimensions	1	1	5	1
	2	4		1
	3	8		
	> 3	13	1	1

IRT more often used for unidimensional data

university of groningen

Exploratory vs confirmatory



university of groningen

Software use

	Type of applied analysis			
	FA ($n = 32$)		IRT	FA & IRT
Software	EFA	CFA	($n = 6$)	($n = 3$)
LISREL		12		1
AMOS		4		
EQS		2		
MPLUS		2		
SCA		1		
NOHARM				1
MSP			2	
RSP				1
TESTGRAF			1	
MULTILOG				1
PARSCALE			1	
WINSTEPS			1	
POLY-SIBTEST			1	
EQUATE				1
DFITPS6				1
SAS	1	1		1
SPSS	1			
STATVIEW	1			
SYSTAT	1			
No information	15	2		1

university of groningen

Outline

university of groningen

Model assumptions: FA ($n = 32$)

- 19 studies: no investigation
- 9 studies: investigated properly
- 4 studies: considered to some extent

Model assumptions: FA ($n = 32$)

- 19 studies: no investigation
- 9 studies: investigated properly
  - Item distributions are examined and reported.
  - Adequate methods (robust) are applied.
- 4 studies: considered to some extent

Model assumptions: FA ($n = 32$)

- 19 studies: no investigation
- 9 studies: investigated properly
- 4 studies: considered to some extent
  - Item distributions are not investigated, but robust estimators used.
  - Both robust and nonrobust analyses, but only reported nonrobust because of similar parameter estimates.

Model assumptions: IRT ($n = 6$)

- 4 studies: investigated properly
  - Unidimensionality assumption investigated
  - IRFs examined for monotonicity
  - Empirical IRFs compared to estimated IRFs
- 2 studies: no investigation

Model fit: FA

CFA: Model fit tested formally usually with measures such as

- RMSEA, GFI, CFI, TLI (NNFI)

EFA: No formal test, but criteria to determine #factors and assignment of items to factors:

- loadings $>$ 0.30 or 0.40
- # factors determined by screeplot, parallel analysis, eigenvalue $>$ 1
- in merely 5 (of 21) studies: interpretability as criterion

Model fit: FA

CFA: Model fit tested formally usually with measures such as

- RMSEA, GFI, CFI, TLI (NNFI)

EFA: No formal test, but criteria to determine #factors and assignment of items to factors:

- loadings $>$ 0.30 or 0.40
- # factors determined by screeplot, parallel analysis, eigenvalue $>$ 1
- in merely 5 (of 21) studies: interpretability as criterion

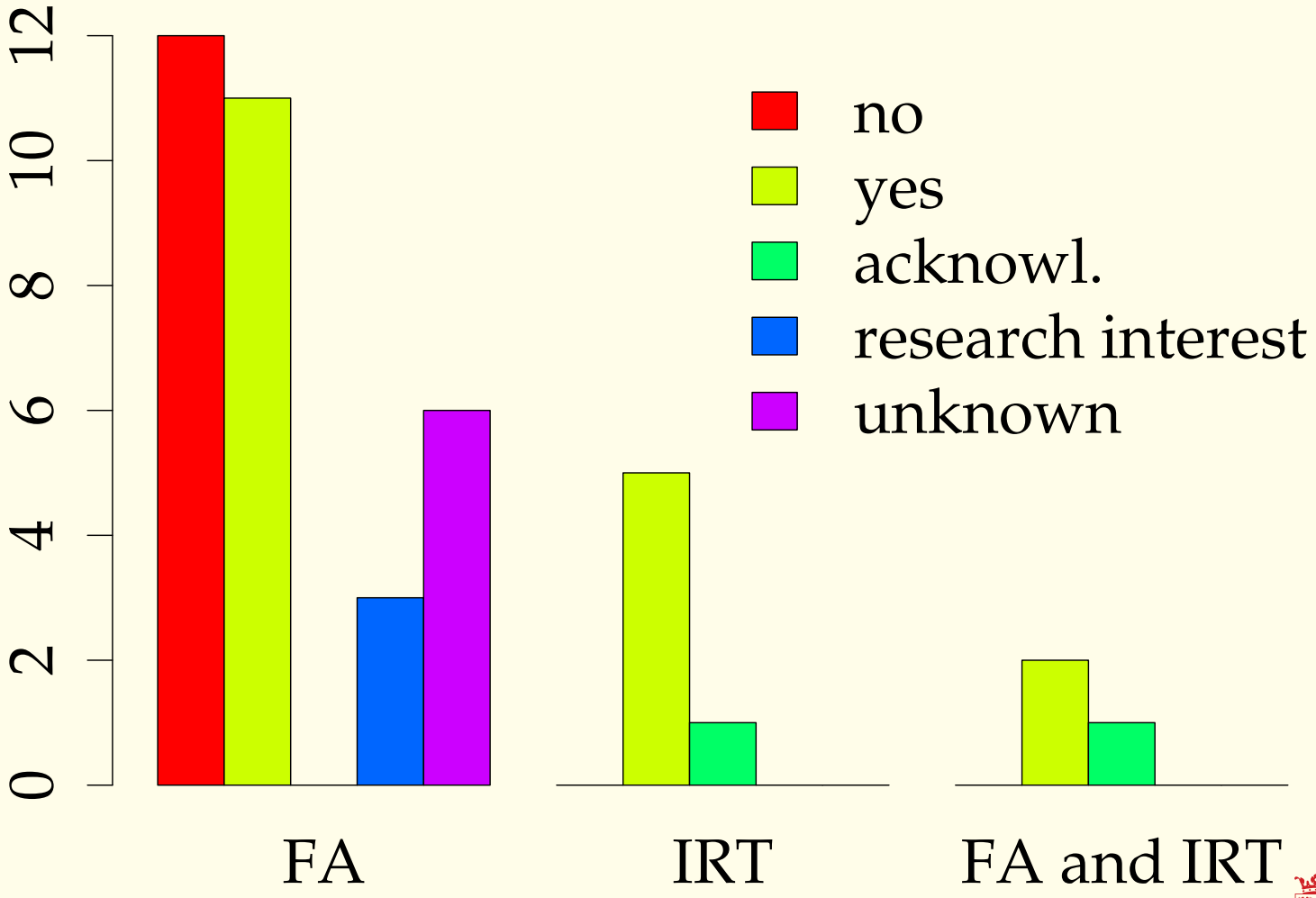
Model fit: IRT ($n = 6$)

- No formal tests reported
- Mokken analysis: Loevinger's H for scale strength
- Unidimensionality tested in 3 studies

Model fit: IRT ($n = 6$)

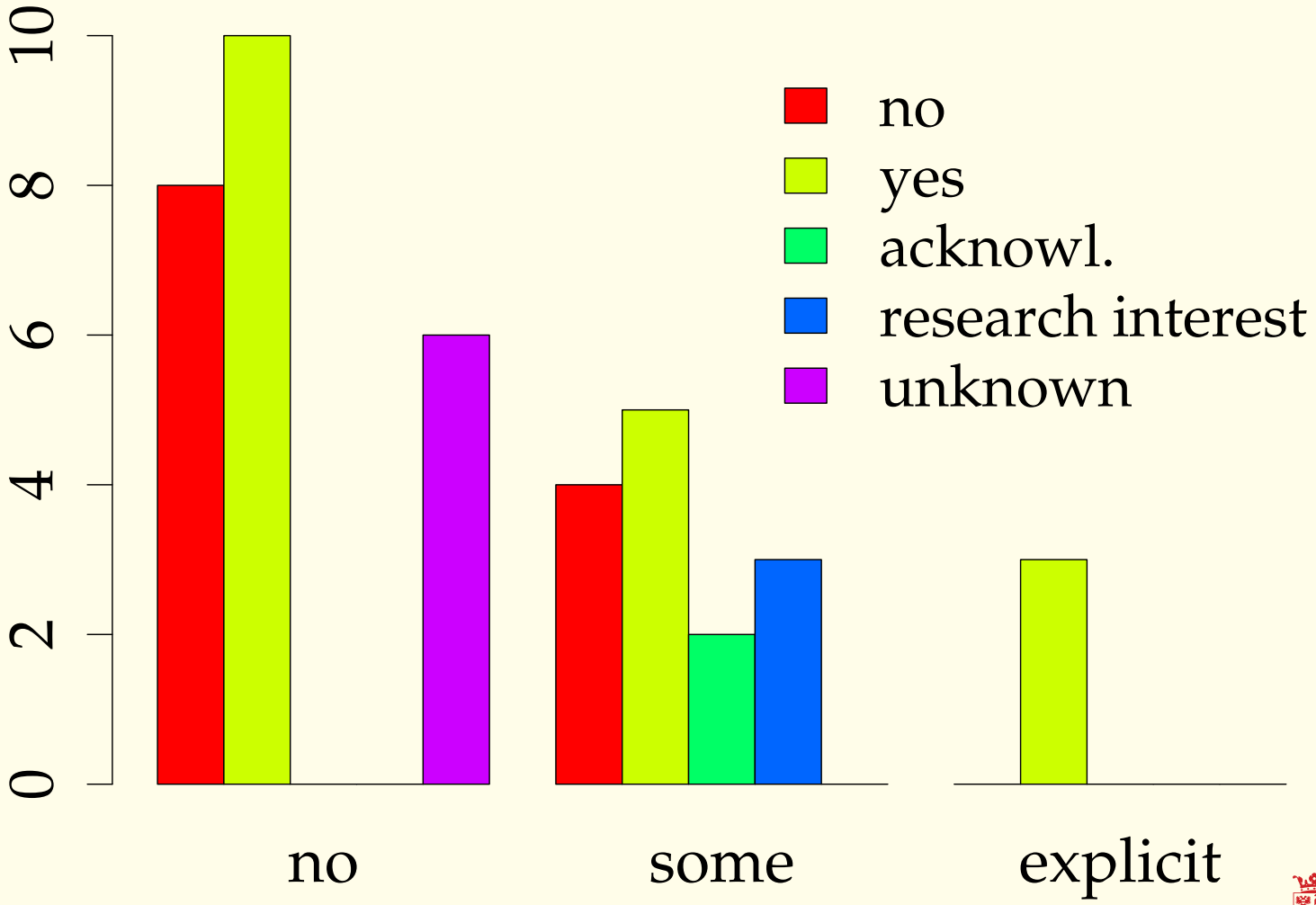
- No formal tests reported
- Mokken analysis: Loevinger's H for scale strength
- Unidimensionality tested in 3 studies

Methodological expert as co-author



university of groningen

Methodological expert as co-author: Motives



university of groningen

Outline

university of groningen

Summary

- FA applied far more often than IRT
- Little explicit motivation in studies
- Possible implicit motives:
  - Expectations about dimensionality
  - FA is more accessible

Recommendations

- Researchers can take better advantage of their theories:
  - More frequent application of confirmatory techniques.
    When applying an exploratory model $\rightarrow$ cross-validate.
  - Add interpretability of factors and content of items to criteria of model evaluation.
- Evaluate model assumptions and report in the paper or on a website.

Future research

- Both simulated and empirical comparisons of FA and IRT
  - Examine impact of violation of model assumptions
  - Extend past research by including nonparametric IRT in the comparison
- Examine differences between latent variable (factor) scores produced by different types of models
- Examine how to combine exploratory and confirmatory approaches in FA and IRT

THANK YOU FOR YOUR ATTENTION

Any questions?

university of groningen

THANK YOU FOR YOUR ATTENTION

Any questions?

university of groningen