Assessing Fit of IRT Models

Nina van Gerwen & Dave Hessen

14th of December, 2022

 $\label{lem:methodology & Statistics for the Behavioural, Biomedical and Social Sciences} \\$





Outline

- Background information
- The issue
- The current study
 - Research questions
- Simulation study
 - Data generation
 - Simulation design
 - Performance metrics
- Sesults preview
- Questions
- References



1. Background information

- Item Response Theory
 - Latent constructs
 - Tests/questionnaires
- Assessing model fit
 - Goodness-of-fit tests
 - Fit indices







What's the issue?

- Few goodness-of-fit tests
 - Multiple issues with current tests ¹
- Scarce studies for TLI² and CFI³



¹(Curran, West, & Finch, 1996; Barton & Lord, 1981)

²(Tucker & Lewis, 1973)

³(Bentler, 1990)

- New goodness-of-fit test
- TLI and CFI

3.1. Research questions



- New goodness-of-fit test
- TLI and CFI

3.1. Research questions

• Required sample size?



- New goodness-of-fit test
- TLI and CFI

3.1. Research questions

- Required sample size?
- 2 Comparing performance



- New goodness-of-fit test
- TLI and CFI

3.1. Research questions

- Required sample size?
- Comparing performance
- Performance of TLI and CFI?



How will we research this?

Through a simulation study of course!





4. Simulation study

4.1. Data generation

- Dichotomous items
- Static item parameters
- Varying four factors





4.2. Simulation Design

Table: Overview of Simulation Conditions for Each Factor

Factor	Conditions	Description		Description	
Test length	5 - 10 - 20	The total number of items that the test will consist of			
Sample size	100 - 200 - 500 1000 - 1500	The total number of observations that are available for each item			
Model type	2PL - 3PL	The models that we will use as the basis for data generation			
Number of groups	2 - 3 - 4	The number of groups that the data gets divided into for the LR Randomisation test calculations			

4.3. Performance metrics

Goodness-of-fit tests

- Power
- ullet Empirical lpha

Fit indices

Mean (SE)





5. Preview of the results

Table 2. Temporary Results for Empirical Alpha

Con	Conditions Goodness-of-fit test					
-	N	LR2		LR4	χ^2	$P-\chi^2$
	IV	LK2	LKS	LK4	χ^{-}	Ρ-χ-
5	100					
	200					
	500					
	1000					
	1500					
10	100					
	200					
	500					
	1000					
	1500					
20	100					
	200					
	500					
	1000					
	1500					

Note. Fitted model = two-parameter logistic model; Data generating model = two-parameter logistic model; I = test length; N = sample size; LR2 = LR Randomisation test with g = 2; LR3 = LR Randomisation test with g = 3; LR4 = LR Randomisation test with g = 4; χ^2 = χ^2 – difference test under the three-parameter logistic model with no constraints; $P_2\lambda^2$ = Pearson's χ^2 test.

Table 4. Fit indices values for correct model specification

Conditions		TLI	CFI	
I	N	M (SE)	M (SE)	
5	100			
	200			
	500			
	1000			
	1500			
10	100			
	200			
	500			
	1000			
	1500			
20	100			
	200			
	500			
	1000			
	1500			

Note. Fitted model = two-parameter logistic model; Data generating model = two-parameter logistic model; I = test length; N = sample size; TLI = tucker-lewis index; CFI = comparative fit index; M = mean; SE = standard error.



Thank you for listening!





References

- Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. ETS Research Report Series, 1981(1), i–8.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, 107(2), 238.
- Curran, P., West, S., & Finch, J. (1996, March). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29. doi: 10.1037/1082-989X.1.1.16
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.

