Studies about Bi-Factor Model are focused on model fit, parameters impacts and other specific analysis. Model fit obliviously is crucial and prevalent topic with many outcomes. Frank Rijmen, (2010) compared bi-factor model with the testlet model and the second-order model with ,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Problem | Model | Analysis | Data | Reference |
| Model fit | Bi-Factor Models ,  Testlet model (2PL) (constrained bi-factor),  Second-order model(constrained bi-factor) | FIML(full-information maximum likelihood)  EM-Newton-Raphson  Deviance  AIC  BIC | testlet-based international English assessment(13,508) | Frank Rijmen (2010) |
| Model fit | Bi-Factor Models,  Correlated Factors, Higher-Order | confirmatory factor analytic (CFA)  ­*X*2  CFI, TLI, RMSEA, AIC, BIC, and aBIC | Monte Carlo Simulations (bi-factor) | Grant B. Morgan  (2015) |
| **Violation of normality**  the effect of skewed trait distributions  Item/population parameter estimation | MGRM (multidimentional graded response model)  IRT  CCFA  UIRT(unidimensional)  MIRT(multimensional) | full-information maximum likelihood (FIML)  robust weighted least  squares (WLS) | Monte Carlo Simulations | Chun Wang  (2018) |
| Model fit | F IRT  UIRT(unidimensional)  MIRT(multimensional) | Type I error  S - *X*2  test length,  sample size, latent trait characteristics  model type misspecification  likelihood  AIC,BIC  Q1 - *X*2  Q1 - *G*2 | Monte Carlo Simulations | Ying Li (2011) |
| vertical scaling(unidimensional)  Several future research points | Bifactor  UIRT  a common vertical scale | Item parameter  Person Para.  Group Para.  ANOVA | Simulation  &  the 2006 fall Michigan mathematics assessments a (common vertical scale) | Ying Li (2012) |
| Model fit  Specific and general facets of hoarding | Bifactor(4factor)  1factor  3factor  4factor  2-order | Y–B *X*2  Delta *X*2  CFI  RMSEA  SRMR | 490 outpatients presenting for psychological treatment at FSU ABHC | modelAmanda M. Raines (2015) |
| **important statistical indices from bifactor models** | quality of unit-weighted total and subscale  score composites  factor score estimates specification quality of a measurement model in (SEM). | SIR(saving inventory revised)  BDI-II(Beck depression inventory-II)  Omega Hierarchical  Contruct reliability  Coefficient Alpha | the standardized factor loadings  from a bifactor model reported in Osman (2009) | Anthony Rodriguez  (2015) |
| Model fit | the single (global) factor model  model with a global factor and a single local factor  a model with a global factor and two local factors  bi-factor model with a global factor and three local factors | BIC  p-ALMR  p-BLRT  *X*2  RMSEA | Monte Carlo Simulations | Tenko Raykov  (2019) |
| **Sample Size Requirements** | confirmatory factor  analysis model | convergence rates, parameter and standard error  estimates ECV | Monte Carlo Simulations | Martina Bader(2022) |

(sample size), (comparison between Bi-factor model and other IRT models, such as)

It is not uncommon for a standardized test to consist of item bundles or testlets

(Bradlow, Wainer, & Wang, 1999); clusters of items that are based on a common

stimulus. Three such multidimensional IRT models are described in the following sections: the

bi-factor model (Gibbons & Hedeker, 1992), the testlet model (Bradlow et al., 1999;

Wainer, Bradlow, & Wang, 2007), and a second-order model.

the formal relation between IRT and factor analysis has

been well established for more than two decades (McDonald, 1982, 1999; Takane &

De Leeuw, 1987).

the estimation method could be extended to accommodate models that relax the proportionality constraint of the testlet model for some but not all testlets.

2

Bi-factor confirmatory factor models have been influential in research on

cognitive abilities because they often better fit the data than correlated factors and

higher-order models.

1]. Subsequently,

the bi-factor model has been recommended by Reise [25] for CFA and successfully employed in the

measurement of a variety of constructs, such as cognitive ability [22], health outcomes [26], quality of

life [27], psychiatric distress [28], early academic skills [29], personality [30], psychopathology [31],

and emotional risk [32].\

As nested models, fit comparisons via ­2 difference tests are possible.

selected 1000 random samples from each population

For example, for each data set we generated from,

say, a population with a true underlying bi-factor model, we fitted a bi-factor, higher-order, and correlated factors CFA model to determine which one(s) fit best.

maximum of 10,000 iterations, Two sample sizes (N = 200, 800) .

a fully crossed design with 36 cells (3 true models 3 fitted models 2 sample sizes 2 factor identification conditions = 36 cells). 1000 replications were run for each cell.

larger. BIC tended to identify the higher-order solution regardless of sample size. When samples were selected from a population with a true multiple correlated factors structure, CFI, TLI, and RMSEA were more likely to identify the correlated factors solution as the best fitting out of the three competing solutions.

When samples were generated from a population with a true higher-order structure, each of the fit indices tended to identify the bi-factor solution as best fitting instead of the true higher-order model. The SRMR had the strongest tendency to prefer the bi-factor model,

3

manipulated: (1) number

of non-normal dimensions (four levels): 0, 1, 2, and 3;

(2) test length (two levels): 30 and 90 items; (3) sample

size (two levels): 500 and 2, 000; (4) correlation between

dimensions (two levels): 0.5 and 0.7.

4 Ying Li (2011)

In recent years, full-information bifactor (FI-bifactor) models have been applied to

empirical data from achievement tests to multiple-domain survey instruments and

have showed superior relative model fit compared with unidimensional item response

theory (UIRT) and multidimensional item response theory (MIRT) models

For the FI-bifactor model data generation conditions, scores on the latent variables

were generated from standard normal distributions under the 2 (sample sizes) × 2

(test lengths) × 3 (levels of item discriminations or factor slopes) ¼ 12 simulation

conditions. For all conditions, 100 replications were used in alignment with previously

published studies to compare the results in this study with previous results at the same

level of Monte Carlo error.

The main purpose of this study was to investigate the performance of the multidimensional

extension of the S w2

statistic in the context of FI-bifactor models in terms of

Type I error rate and power. It was found that nominal Type I error rates for FI-bifactor

models were observed for most conditions and were not influenced by test length,

sample size, or loading structures. Therefore, it was concluded that the central chisquare

distribution under the null hypothesis approximated the sampling distribution

of the S w2

statistic for FI-bifactor models reasonably well in the tail.

This study was the first that examined the power of the S w2

statistic for UIRT

models to detect model–data misfit when data were generated using multidimensional

models, which complements the results of Orlando and Thissen (2000).

5YingLi(2012)

Bifactor model estimation accuracy was evaluated through a

simulation study with manipulated factors of percentage of common items, sample size, and

degree of construct shift.

Limitations and Directions for Future Research

This study examined only one of the three data collection designs (i.e., common-item design) for vertical scaling. Investigations can be extended to the performance of the bifactor model vertical scaling for the equivalent groups design and the scaling test design. In terms of item type, the current study considered tests with only dichotomously scored

items. Future studies can be extended to polytomously scored items or even mixed item format

tests. In terms of the bifactor item response function (Rijmen, 2010), this study considered a

two-parameter (difficulty and discrimination) bifactor model; examination of a three-parameter

(difficulty, discrimination, and guessing parameters) bifactor model or simplification to a oneparameter

(difficulty parameter only) bifactor model can also be conducted.