A new reliability coefficient based on latent class analysis

L. Andries van der Ark & Daniël W. van der Palm
Tilburg University

FIGURE 6.1 The Rosenberg Self-Esteem Scale

Circle one response for each of the following ten items.

	Strongly Agree	*Agree*	*Disagree*	*Strongly Disagree*
1. I feel that I am a person of worth, at least on an equal basis with others.	1	2	3	4
2. I feel that I have a number of good qualities.	1	2	3	4
*3. All in all, I am inclined to feel that I am a failure.	1	2	3	4
4. I am able to do things as well as most other people.	1	2	3	4
*5. I feel I do not have much to be proud of.	1	2	3	4
6. I take a positive attitude toward myself.	1	2	3	4
7. On the whole, I am satisfied with myself.	1	2	3	4
*8. I wish I could have more respect for myself.	1	2	3	4
*9. I certainly feel useless at times.	1	2	3	4
*10. At times I think I am no good at all.	1	2	3	4

*Items marked with an asterisk have reversed wording. The numbers on items with reversed wording should be reversed before summing the responses for the ten items. For example, on item 3, "strongly agree" becomes 4, "agree" becomes 3, "disagree" becomes 2, and "strongly disagree" becomes 1.

Source: Morris Rosenberg's "Self-Esteem Scale" from pp. 325–327 of *Society and Adolescent Self-Image*

Motivation

- Development of R package for *Mokken scale analysis*
  `mokken`: http://cran.r-project.org/web/packages/mokken
- Reliability most reported coefficient in test construction
  Mokken scale analysis: MS statistic rather than α.
- Research questions

  1. Compare bias and accuracy of well-known reliability coefficients

  2. Propose a new reliability coefficient based on latent class analysis.

Test: J items (i, j) with ordered categories $0, \ldots, m$ (x, y)

$$X = \sum_{j=1}^{J} X_j$$

Classical test theory

$$\underbrace{X}_{\text{Test score}} = \underbrace{T}_{\text{True score}} + \underbrace{E}_{\text{Error}}$$

Reliability

$$r_{XX'} = \frac{S^2(T)}{S^2(X)}$$

Reliability estimation: Find estimate of $S^2(T)$

Table of Marginal and Joint Cumulative Probabilities

Margins: $P_{x(i)} = P(X_i \geq x)$

Cells: $P_{x(i),y(j)} = P(X_i \geq x, X_j \geq y)$ (48 observed, 16 unobserved)

	Unpopular							Popular
	$P_{2(4)}$	$P_{2(3)}$	$P_{2(2)}$	$P_{2(1)}$	$P_{1(4)}$	$P_{1(3)}$	$P_{1(2)}$	$P_{1(1)}$
$P_{2(4)}$	$P_{2(4),2(4)}$	$P_{2(4),2(3)}$	$P_{2(4),2(2)}$	$P_{2(4),2(1)}$	$P_{2(4),1(4)}$	$P_{2(4),1(3)}$	$P_{2(4),1(2)}$	$P_{2(4),1(1)}$
$P_{2(3)}$	$P_{2(3),2(4)}$	$P_{2(3),2(3)}$	$P_{2(3),2(2)}$	$P_{2(3),2(1)}$	$P_{2(3),1(4)}$	$P_{2(3),1(3)}$	$P_{2(3),1(2)}$	$P_{2(3),1(1)}$
$P_{2(2)}$	$P_{2(2),2(4)}$	$P_{2(2),2(3)}$	$P_{2(2),2(2)}$	$P_{2(2),2(1)}$	$P_{2(2),1(4)}$	$P_{2(2),1(3)}$	$P_{2(2),1(2)}$	$P_{2(2),1(1)}$
$P_{2(1)}$	$P_{2(1),2(4)}$	$P_{2(1),2(3)}$	$P_{2(1),2(2)}$	$P_{2(1),2(1)}$	$P_{2(1),1(4)}$	$P_{2(1),1(3)}$	$P_{2(1),1(2)}$	$P_{2(1),1(1)}$
$P_{1(4)}$	$P_{1(4),2(4)}$	$P_{1(4),2(3)}$	$P_{1(4),2(2)}$	$P_{1(4),2(1)}$	$P_{1(4),1(4)}$	$P_{1(4),1(3)}$	$P_{1(4),1(2)}$	$P_{1(4),1(1)}$
$P_{1(3)}$	$P_{1(3),2(4)}$	$P_{1(3),2(3)}$	$P_{1(3),2(2)}$	$P_{1(3),2(1)}$	$P_{1(3),1(4)}$	$P_{1(3),1(3)}$	$P_{1(3),1(2)}$	$P_{1(3),1(1)}$
$P_{1(2)}$	$P_{1(2),2(4)}$	$P_{1(2),2(3)}$	$P_{1(2),2(2)}$	$P_{1(2),2(1)}$	$P_{1(2),1(4)}$	$P_{1(2),1(3)}$	$P_{1(2),1(2)}$	$P_{1(2),1(1)}$
$P_{1(1)}$	$P_{1(1),2(4)}$	$P_{1(1),2(3)}$	$P_{1(1),2(2)}$	$P_{1(1),2(1)}$	$P_{1(1),1(4)}$	$P_{1(1),1(3)}$	$P_{1(1),1(2)}$	$P_{1(1),1(1)}$

$$r_{XX'} = \frac{S^2(T)}{S^2(X)} = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right] + \sum_{i=j} \left[P_{x(i),y(i)} - P_{x(i)}P_{y(i)}\right]}{S^2(X)}$$

1. **Cronbach's alpha** (α; Cronbach, 1950)

$$r_{XX'} = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right] + \sum_{i=j} \left[P_{x(i),y(i)} - P_{x(i)}P_{y(i)}\right]}{S^2(X)}$$

$$\alpha = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right] + \sum_{i=j} \left[\overline{P_{xy} - P_xP_y}\right]}{S^2(X)}$$

- $\overline{P_{xy} - P_xP_y}$ mean of all observed $P_{x(i),y(j)} - P_{x(i)}P_{y(j)}$s
- Most popular estimator of reliability
- α lower bound to reliability ($\alpha \leq r_{XX'}$)
- Much literature on se(α)

2. Lambda 2 (λ_2; Guttman, 1945)

$$r_{XX'} = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right] + \sum_{i=j} \left[P_{x(i),y(i)} - P_{x(i)}P_{y(i)}\right]}{S^2(X)}$$

$$\lambda_2 = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right] + \sqrt{\frac{J}{J-1} \sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)}P_{y(j)}\right]^2}}{S^2(X)}$$

- Unpopular estimator of reliability
- Better lower bound: $\alpha \leq \lambda_2 \leq r_{XX'}$
- λ_2 not the greatest lower bound (GLB)

3. MS (Molenaar & Sijtsma, 1984; 1988)

$$r_{XX'} = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)} P_{y(j)}\right] + \sum_{i=j} \left[P_{x(i),y(i)} - P_{x(i)} P_{y(i)}\right]}{S^2(X)}$$

$$MS = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)} P_{y(j)}\right] + \sum_{i=j} \left[\overline{P}_{x(i),y(i)} - P_{x(i)} P_{y(i)}\right]}{S^2(X)}$$

- Direct estimator
- $P_{x(i),y(i)}$ estimated in 8 different ways:
  Their mean $\overline{P}_{x(i),y(i)}$ is used.
- Based on assumptions that ISRFs do not intersect and uni-dimensionality.

4. Latent class (LC) based reliability statistic (LCR)

$$r_{XX'} = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)} P_{y(j)} \right] + \sum_{i=j} \left[P_{x(i),y(i)} - P_{x(i)} P_{y(i)} \right]}{S^2(X)}$$

$$LCR = \frac{\sum_{i \neq j} \left[P_{x(i),y(j)} - P_{x(i)} P_{y(j)} \right] + \sum_{i=j} \left[\widehat{P}_{x(i),y(i)} - P_{x(i)} P_{y(i)} \right]}{S^2(X)}$$

- $\widehat{P}_{x(i),y(i)}$: LC estimates of $P_{x(i),y(i)}$.
- Idea: Fit LC model to data and derive $\widehat{P}_{x(i),y(i)}$ from LC parameters.

Derivation of $\widehat{P}_{x(i),y(i)}$ from LC parameters.

LC model:

> Each respondent is a member of one (out of K) unobservable classes. Within each class all item scores are independent.

$$
\begin{aligned}
P(X_i = x, X_j = y | T = k) &= P(X_i = x | T = k) \times P(X_j = y | T = k) \\
P(X_i = x, X_i = y | T = k) &= P(X_i = x | T = k) \times P(X_i = y | T = k) \Rightarrow \\
P(X_i = x, X_i = y) &= \sum_{k=1}^{K} P(T = k) \times P(X_i = x | T = k) \times P(X_i = y | T = k) \\
P(X_i \geq x, X_i \geq y) &= \sum_{u=x}^{m} \sum_{v=y}^{m} \sum_{k=1}^{K} P(T = k) \times P(X_i = u | T = k) \times P(X_i = v | T = k) \\
\widehat{P}(X_i \geq x, X_i \geq y) &= \sum_{u=x}^{m} \sum_{v=y}^{m} \sum_{k=1}^{K} \widehat{P}(T = k) \times \widehat{P}(X_i = u | T = k) \times \widehat{P}(X_i = v | T = k)
\end{aligned}
$$

4. Latent class (LC) based reliability statistic (LCR)

- Except local independence, no data assumptions

- Latent class model used for estimation of distribution only.

- Main question: Does the LC model fit? (number of classes, other fit issues)

Expectations:

1. LCR unbiased if model fits (BIC)
2. α and λ_2 negatively biased
3. MS unbiased only if unidimensionality holds and ISRFs don't intersect

Simulation study- Method:

1. GRM ($\theta \sim N(0, 1)$) was used to compute population reliability and to generate data.
2. Full experimental design (1000 replications)
3. Dependent variables: *Bias* and *accuracy* of coefficients
4. Independent variables: *Coefficients*, *dimensionality*, *item format*, *test length*, *item discrimination*, *item location*, *sample size*, *number of latent classes*

Simulation study- Results/Discussion:

- **LCR: Number of latent classes required.**
- Best results LCR: Number of classes in LCM with lowest BIC +1.
  Less classes: negative bias
  More classes: positive bias

- Fit: compare $P(X_i = x, X_j = y)$ with $\sum_k \widehat{P}(T = k)\widehat{P}(X_i = x|T = k)\widehat{P}(X_j = y|T = k)$.
- Not $P(X_i = x, X_i = y)$ with $\sum_k \widehat{P}(T = k)\widehat{P}(X_i = x|T = k)\widehat{P}(X_i = y|T = k)$.
- Problematic for

  1. Wrong number of latent classes

$$P(X_j = x) = \sum_{k=1}^{K} \widehat{P}(T = k)\widehat{P}(X_j = x|T = k) \approx \sum_{k=1}^{K+1} \widehat{P}(T^* = k)\widehat{P}(X_j = x|T^* = k)$$

but

$$P(X_j = x, X_j = x) = \sum_{k=1}^{K} \widehat{P}(T = k)\widehat{P}(X_j = x|T = k)^2 < \sum_{k=1}^{K+1} \widehat{P}(T^* = k)\widehat{P}(X_j = x|T^* = k)^2$$

Simulation study- Results/Discussion:
Unidimensional tests.

- Accuracy similar over coefficients
- Negative bias for α $[-.03; -.08]$ and λ_2 $[-.01; -.04]$
- MS almost unbiased when ISRFs do not intersect $[-.005; +.005]$
  MS very slightly biased when ISRFs intersect $[-.015; +.010]$
  Computation time fast ($<$ 1 second)
- LCR almost unbiased if model fits $[-.003; +.003]$
  LCR more biased than MS if too many/few latent classes.
  Computation time reasonable (3 or 4 seconds)
- Practical advise: MS

Simulation study- Results/Discussion:
Multidimensional tests.

- Accuracy similar over coefficients
- Large negative bias for α, λ_2 and MS $[-.05; -.10]$
- LCR (BIC+1) slightly biased $[-.02; +.02]$

- Practical advise: LCR
- To do: Compare with coefficients for heterogeneous tests.