Factor Analytic Approaches to Demonstrating Internal Structure Validity of Test Scores: An Overview of the Major Approaches to Factor Analysis and When to use Them

Mike Kalkbrenner, Ph.D., NCC

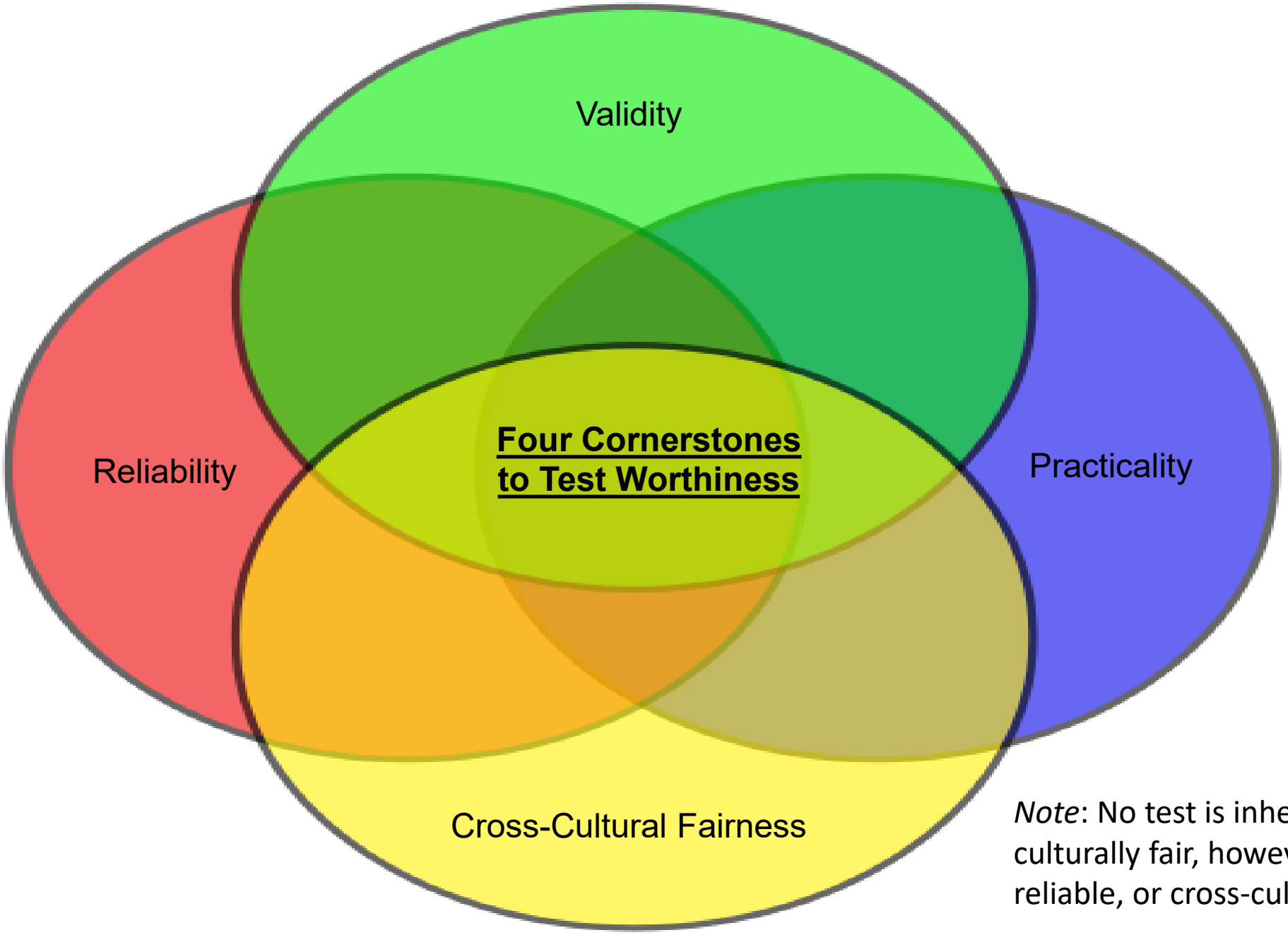
College of Health, Education, and Social Transformation

Department of Counseling and Educational Psychology

BE BOLD. Shape the Future.®
New Mexico State University

Agenda

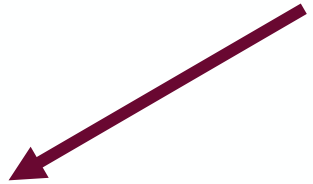
- Welcome and introduction
- Use an existing measure or develop your own?
- Overview of the major approaches factor analysis
- Comments and questions!



Note: No test is inherently valid, reliable, or cross-culturally fair, however, test scores can be valid, reliable, or cross-culturally fair.

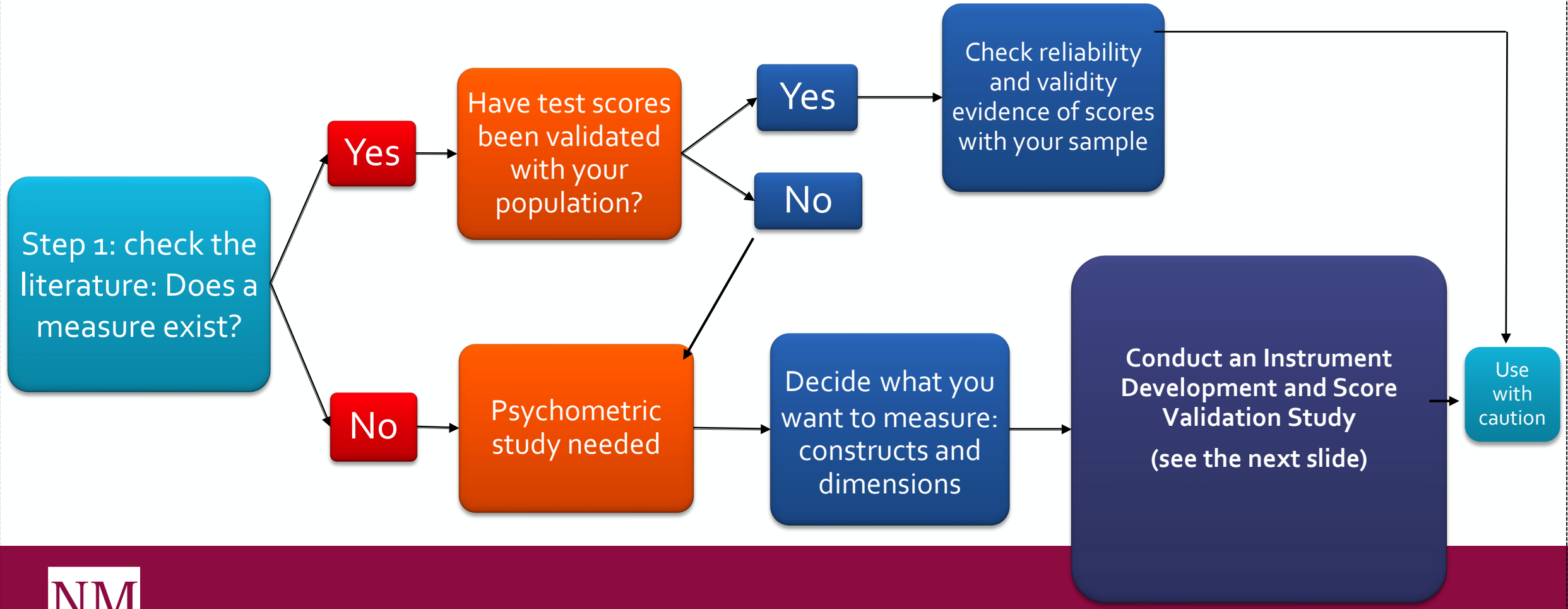
Validity Evidence of Test Scores

- A unitary concept; however, there are 5 general sources of validity evidence of test scores (American Educational Research Association [AERA], 2014):
    1. Content Validity
    2. Response Process Validity
    3. Internal Structure (Factor Analysis)
    4. Relations to Other Variables
        - Convergent and Discriminant Validity
        - Test Criterion Validity
            - Concurrent and Predictive Validity
    5. Consequential Validity



Today's presentation focuses on

To Develop or Not to Develop: Determining whether to use an existing measure from the literature or develop your own



Adapted from Sink (2016)

The MEASURE Approach to Instrument Development

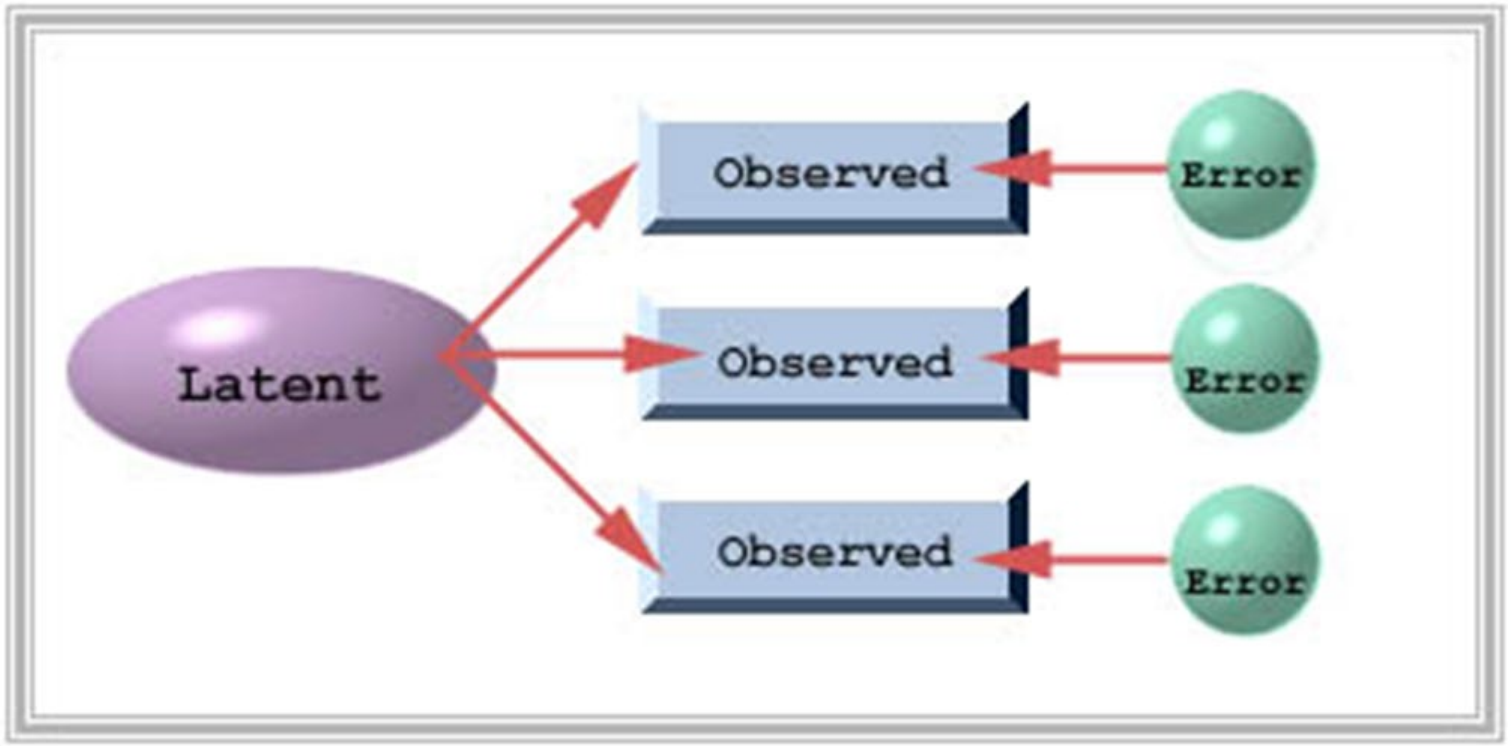
- **M**ake the purpose and rationale clear
- **E**stablish empirical framework
- **A**rticulate theoretical blueprint
- **S**ynthesizing content and scale development
- **U**se expert reviewers
- **R**ecruit participants
- **E**valuate validity and reliability

Kalkbrenner (2021): https://doi.org/10.7275/svg4-e671

What is Factor Analysis?

- A series of statistical analyses for estimating *internal structure* validity of test scores.
- Internal structure validity:
  - The extent to which the overall items on an instrument measure a coherent *latent variable* (i.e., theoretical or hypothetical trait).
    - For example:
      - Intelligence
      - Self-esteem
      - Empathy
      - Classroom climate
      - Health literacy
      - Resilience
      - Student engagement
      - And more!!!

Latent vs. Observed Variables



Major Approaches to Factor Analysis

- Exploratory Factor Analysis (EFA)
- Confirmatory Factor Analysis (CFA)
  - Extensions of Confirmatory Factor Analysis:
    - Higher-Order Confirmatory Factor Analysis
    - Bi-Factor Confirmatory Factor Analysis
    - Multiple-Group Confirmatory Factor Analysis

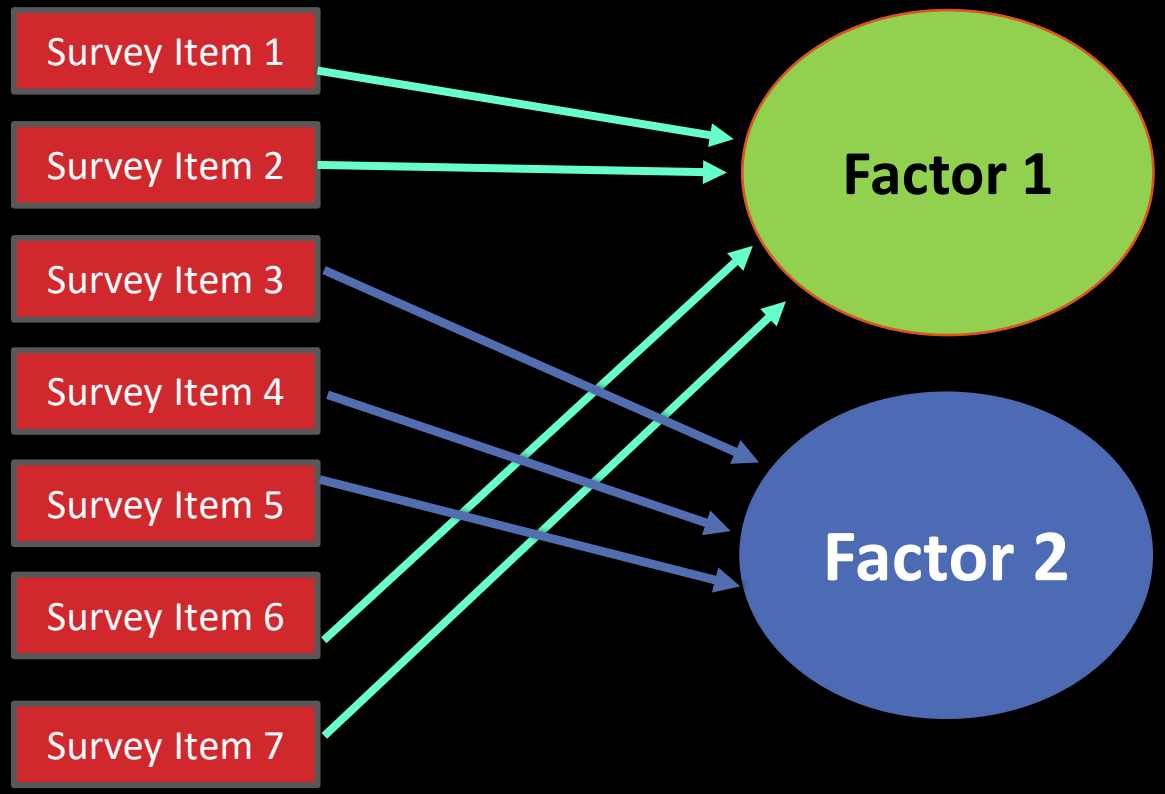
Primary Aim of Exploratory Factor Analysis (EFA)

- To simplify an inter-item correlation matrix in a few underlying dimensions (or factors) that make sense both statistically and logically.
    - Looking for a range of inter-item correlations between approximately .20 to .80 (Why?)

Determining Sample Size for Factor Analysis

- Many guidelines for sample size:
  - 10 participants per parameter to be estimated (Kline, 2005; Byrne, 2019)
  - 20 participants per parameter (Tanaka, 1987)
  - 5 participants per parameter (Bentler & Chou, 1987)
- At least 200 participants regardless of the number of estimated parameters

EXPLORATORY FACTOR ANALYSIS



*In EFA the items lead to the emergence of the factors or subscales

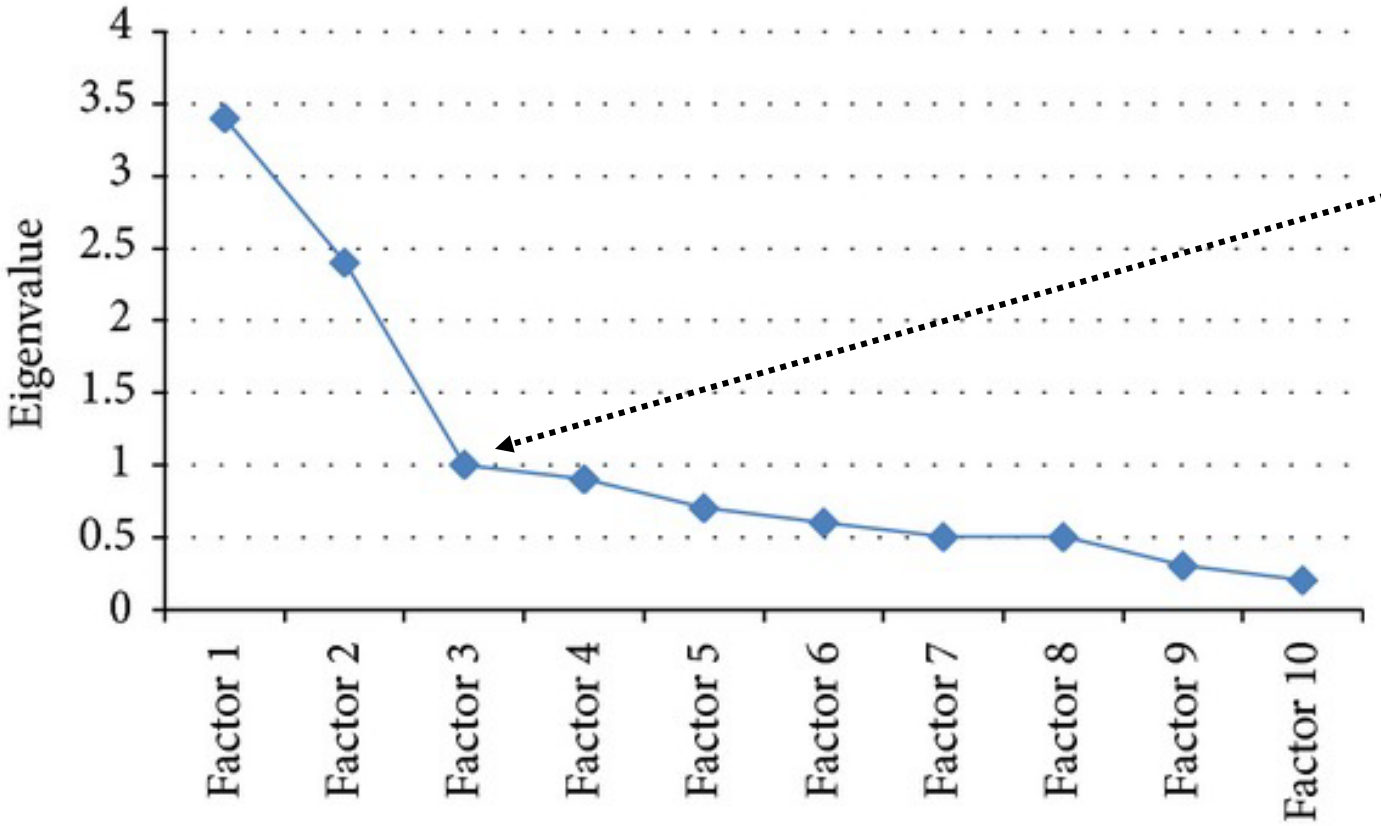
Primary Factor Extraction Methods

- Principal Components Analysis (PCA)
  - A type of factor analysis or a method of data reduction?
- Principal Axis Factoring (PAF)
- Maximum Likelihood (ML)

Determining the Number of Factors to Extract

- Kaiser Criterion
  - Drop all factors with Eigenvalues under 1.0
- Meaningful Variance (more than 5%)
- Cattell's Scree Test
- Parallel Analysis

Example Scree Plot



Indicates a 3-factor solution

Parallel Analysis

- Comparison between the sample data and a matrix of random numbers.
  - Eigenvalues are compared between the sample data and the random data.
    - Retain the number of factors that have larger Eigenvalues compared to the sample data.

Parallel Analysis

3 factors are retained

TABLE 4

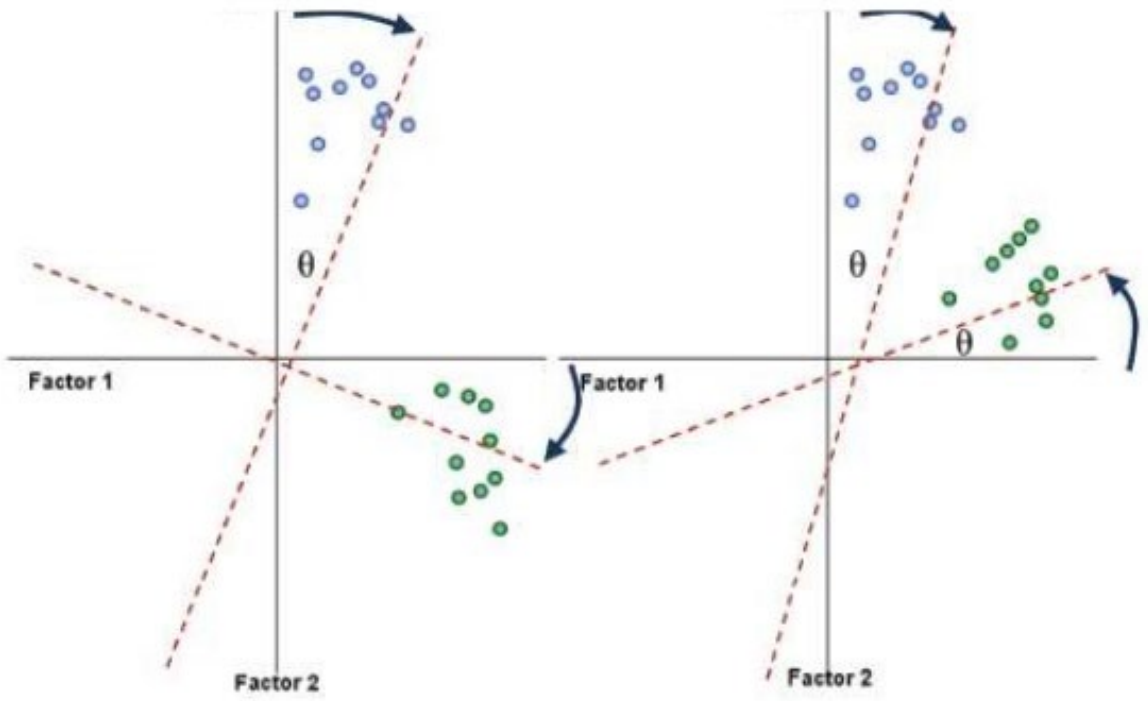
Eigenvalues Derived by Parallel Analysis of Principal Components for the Life Regard Index ($N = 334$)

Real Data	Random Permutations of Real Data
10.37 >	1.58
2.04 >	1.49
1.56 >	1.43
1.23	1.37
1.10	1.33
0.90	1.28
0.85	1.24

(Steger, 2007)

Factor Rotation

1) Orthogonal (most commonly varimax)
   - Rotate the data on vectors at 90-degree angles
2) Oblique Rotation
   - Rotate data at angles that are less than 90-degrees



Orthogonal Rotation Oblique Rotation

Extracted Communalities

- Indicates % of variance in each observed variable (test item) that each factor explains
- Higher communality (or h^2), more the component or factor will explain of the variance of each item or variable.
- h^2 values should be $\geq .30$
- Re-compute the EFA after removing each item

Communalities

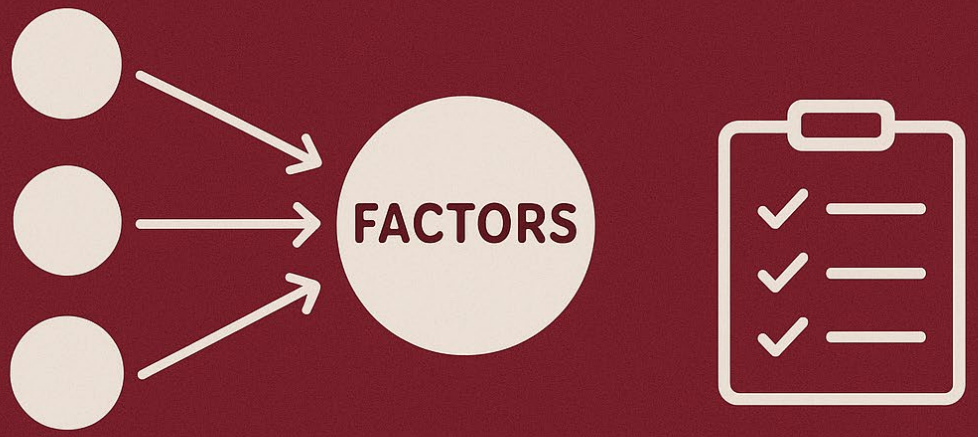
	Initial	Extraction
behav1 CONCENTRATES	.713	.746
behav2 CURIOUS	.743	.788
behav3 PERSEVERES	.766	.811
behav4 EVEN-TEMPERED	.729	.747
behav5 PLACID	.609	.664
behav6 COMPLIANT	.687	.710
behav7 SELF-CONTROLLED	.730	.749
behav8 RELATES-WARMLY	.605	.660
behav9 SUSTAINED ATTENTION	.776	.803
behav10 COMMUNICATIVE	.657	.674
behav11 RELAXED	.786	.820
behav12 CALM	.737	.786
behav13 PURPOSEFUL ACTIVITY	.764	.798
behav14 COOPERATIVE	.626	.647
behav15 CONTENTED	.595	.621

Extraction Method: Principal Axis Factoring.

Factor Loadings

- Values that denote the strength of relationship between observed variables (i.e., items) and the latent factor.
- Tentative guidelines for interpreting factor loadings:
  - "Weak" if less than .39
  - "Moderate" between .40 - .59
  - "Strong" if more than .60
  - Cross-loading if more than .35 on two or more factors
  - Generally, the minimum cutoff for marking a factor is $\geq$ .40

LET'S PRACTICE!!



EXPLORATORY FACTOR ANALYSIS

Life Satisfaction Survey (Questions)	Factor Loadings			
	1	2	3	4
1. I Feel tired most of the time	.82		.12	
2. I have trouble falling asleep	.73		.25	
3. I have difficulty staying asleep	.68		.14	
4. I enjoy sleeping	.55		.39	.34
5. I feel excited before going to work	.14	.78	.12	
6. My work is meaningful	.26	.70		
7. My supervisor respect my opinions	.23	.66	.24	
8. My coworkers are also my friends		.52		
9. I like having fun	.28		.17	.32
10. I can afford to buy what I need			.61	
11. Money is not a source of stress for me			.59	
12. I can buy a leisure item when I want to			.48	
13. I like money	.29	.37	.41	
14. I feel connected with others around me				.60
15. The people closest to me care about me	.13			.58
16. I can rely on my friends to have my back		.24		.47
17. My friends feel like family	.21			.42

Let's Practice!

*We are looking for items that clearly load on one factor ($\geq$.40) and do not cross-load ($\geq$.35) on two or more factors.

Life Satisfaction Survey (Questions)	Factor Loadings			
	1	2	3	4
1. I Feel tired most of the time	.82		.12	
2. I have trouble falling asleep	.73		.25	
3. I have difficulty staying asleep	.68		.14	
~~4. I enjoy sleeping~~	~~.55~~		~~.39~~	~~.34~~
5. I feel excited before going to work	.14	.78	.12	
6. My work is meaningful	.26	.70		
7. My supervisor respect my opinions	.23	.66	.24	
8. My coworkers are also my friends		.52		
~~9. I like having fun~~	~~.28~~		~~.17~~	~~.32~~
10. I can afford to buy what I need			.61	
11. Money is not a source of stress for me			.59	
12. I can buy a leisure item when I want to			.48	
~~13. I like money~~	~~.29~~	~~.37~~	~~.41~~	
14. I feel connected with others around me				.60
15. The people closest to me care about me	.13			.58
16. I can rely on my friends to have my back		.24		.47
17. My friends feel like family	.21			.42

***Important:** Remove items one at a time and re-compute the EFA.

← Item # 4 cross-loads on factors 1 & 3

← Item # 9 fails to load (≤.40) on any factor

← Item # 13 cross-loads on factors 2 & 3

Life Satisfaction Survey (Questions)	Factor Loadings			
	1	2	3	4
1. I Feel tired most of the time	**.82**		.12	
2. I have trouble falling asleep	**.73**		.25	
3. I have difficulty staying asleep	**.68**		.14	
5. I feel excited before going to work	.14	**.78**	.12	
6. My work is meaningful	.26	**.70**		
7. My supervisor respect my opinions	.23	**.66**	.24	
8. My coworkers are also my friends		**.52**		
10. I can afford to buy what I need			**.61**	
11. Money is not a source of stress for me			**.59**	
12. I can buy a leisure item when I want to			**.48**	
14. I feel connected with others around me				**.60**
15. The people closest to me care about me	.13			**.58**
16. I can rely on my friends to have my back		.24		**.47**
17. My friends feel like family	.21			**.42**

Note: Factor loadings over .40 appear in bold and mark the particular factor. Blank cells indicate factor loadings ≤ .10.

Life Satisfaction Survey (Questions)	Factor Loadings			
	1	2	3	4
1. I Feel tired most of the time	**.82**		.12	
2. I have trouble falling asleep	**.73**		.25	
3. I have difficulty staying asleep	**.68**		.14	
5. I feel excited before going to work	.14	**.78**	.12	
6. My work is meaningful	.26	**.70**		
7. My supervisor respect my opinions	.23	**.66**	.24	
8. My coworkers are also my friends		**.52**		
10. I can afford to buy what I need			**.61**	
11. Money is not a source of stress for me			**.59**	
12. I can buy a leisure item when I want to			**.48**	
14. I feel connected with others around me				**.60**
15. The people closest to me care about me	.13			**.58**
16. I can rely on my friends to have my back		.24		**.47**
17. My friends feel like family	.21			**.42**

POSSIBLE FACTOR NAMES

Factor 1: Sleep Difficulty

Factor 2: Work Satisfaction

Factor 3: Financial Stability

Factor 4: Social Connectedness

CONFIRMATORY FACTOR ANALYSIS

Table 1

Fit Indices and Tentative Thresholds for Evaluating Model Fit

Absolute Fit Indices				
Index	*Abbreviation*	*Strong Fit*	*Acceptable Fit*	*Poor Fit*
Chi-square	χ^2 or CMIN	p-value > .05 or χ^2 to $df \leq 1$	χ^2 to $df \leq 2$ or 3	χ^2 to $df > 3$
Standardized root mean square residual	SRMR	< .05	.06 to .08	> .08
Root mean square error of approximation	RMSEA	< .05, report confidence interval	.06 to .08 (.081 to .10 can denote a somewhat acceptable fit)	> .10
Goodness-of-fit index & Adjusted goodness-of-fit index	GFI/AGFI	≥ .97	≥ .95 (≥ .90 to .94 can denote a somewhat acceptable fit)	< .90
Incremental Fit Indices				
Index	*Abbreviation*	*Strong Fit*	*Acceptable Fit*	*Poor Fit*
Comparative fit index	CFI	≥ .97	.95 to .90	< .90
Normed fit index	NFI	≥ .97	.95 to .90	< .90
Incremental fit index	IFI	≥ .97	.95 to .90*	< .90
Tucker–Lewis index	TLI	≥ .97	.95 to .90	< .90
Parsimonious Fit Indices				
Index	*Abbreviation*	*Strong Fit*	*Acceptable Fit*	*Poor Fit*
Parsimony-adjusted goodness-of-fit index	PGFI	Parsimony-adjusted indices range from 0 to 1 and have utility for making comparisons between different models. Values closer to 1 indicate a stronger fit. (Kalkbrenner, 2021)		
Parsimony-adjusted normed fit index	PNFI			


LET'S GIVE IT A TRY!!
FACTORS
CONFIRMATORY FACTOR ANALYSIS

Model 1

- Evaluate the following model fit in terms of *excellent*, *acceptable*, or *poor*
  - *CMIN* = $\chi 2$ (77) = 200.01, $p < .001$, χ^2 to *df* = 2.60
  - Comparative Fit Index (*CFI*) = .97
  - Root mean square error of approximation (*RMSEA*) = .04, 90% *CI* (.02, .06)
  - Standardized root mean square residual (*SRMR*) = .03

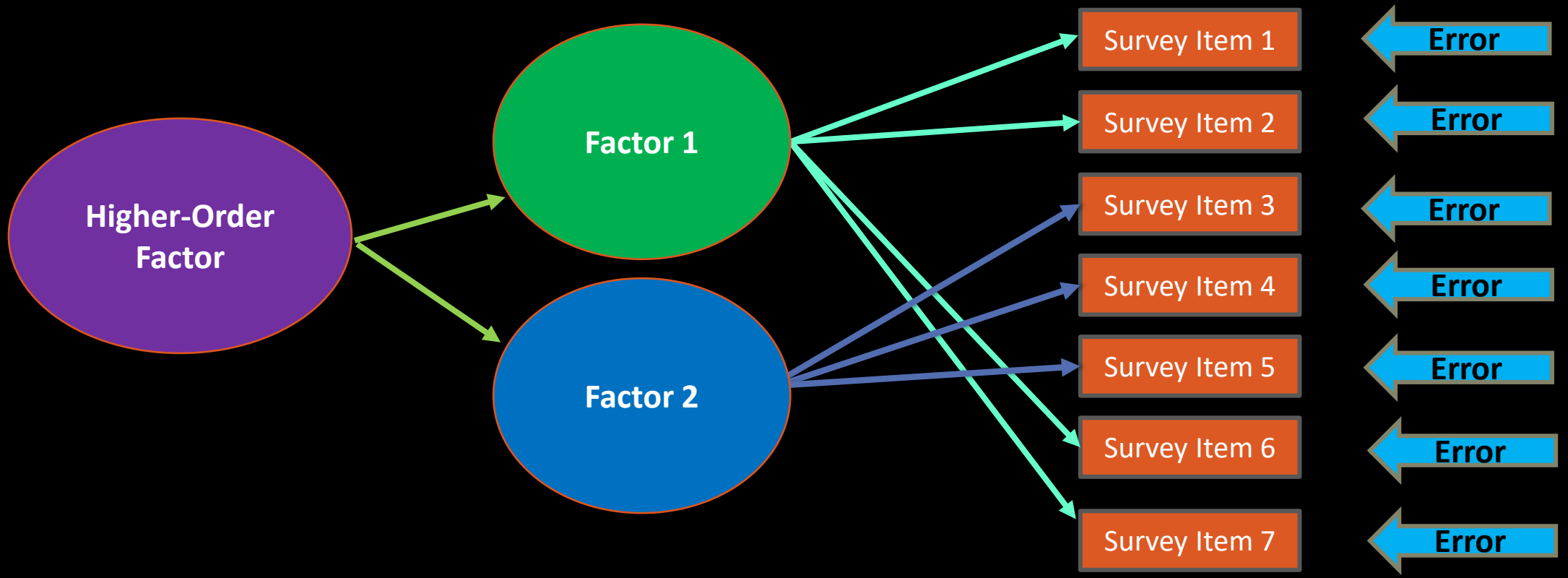
Model 2

- Evaluate the following model fit in terms of *excellent*, *acceptable*, or *poor*
  - *CMIN* = $\chi 2$ (74) = 357.93, $p < .001$, χ^2 to *df* = 4.84
  - Comparative Fit Index (*CFI*) = .90
  - Root mean square error of approximation (*RMSEA*) = .09, 90% *CI* (.08, .10)
  - Standardized root mean square residual (*SRMR*) = .14

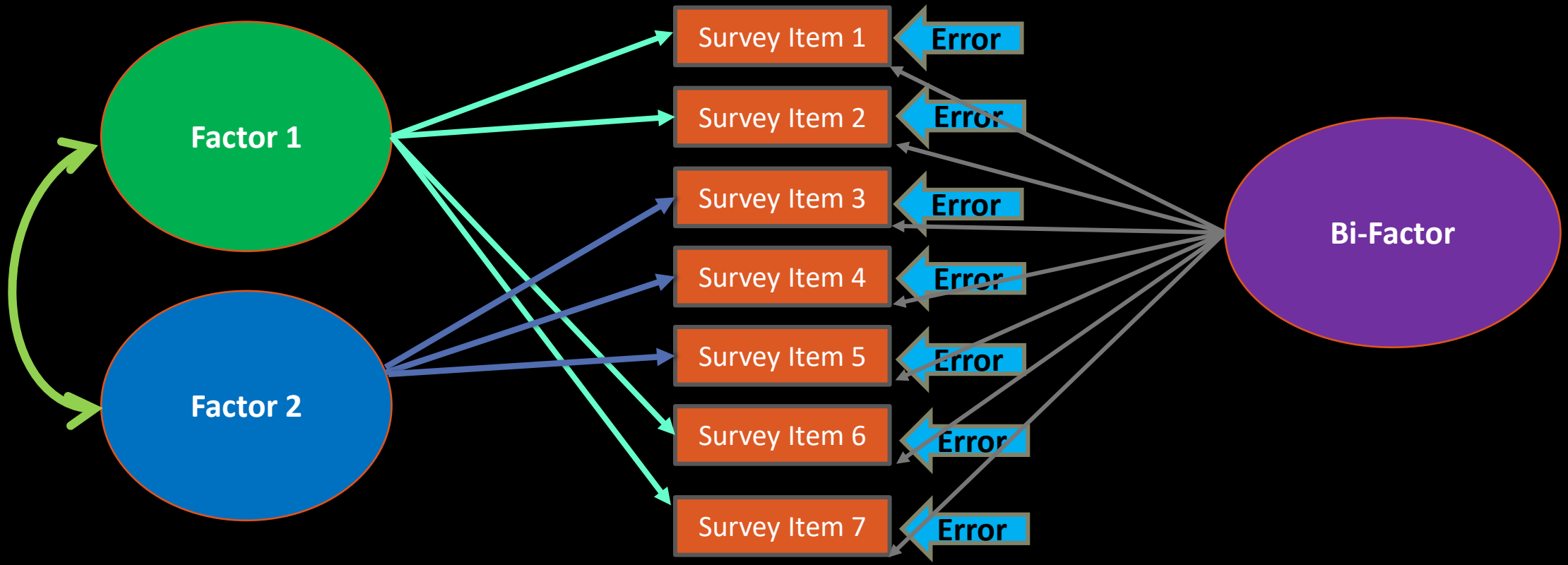
Model 3

- Evaluate the following model fit in terms of *excellent*, *acceptable*, or *poor*
  - *CMIN* = $\chi 2$ (140) = 400.33, $p < .001$, χ^2 to *df* = 2.86
  - Comparative Fit Index (*CFI*) = .91
  - Root mean square error of approximation (*RMSEA*) = .09, 90% *CI* (.07, .12)
  - Standardized root mean square residual (*SRMR*) = .08

HIGHER-ORDER CONFIRMATORY FACTOR ANALYSIS



BI-FACTOR CONFIRMATORY FACTOR ANALYSIS



Multiple-Group Confirmatory Factor Analysis

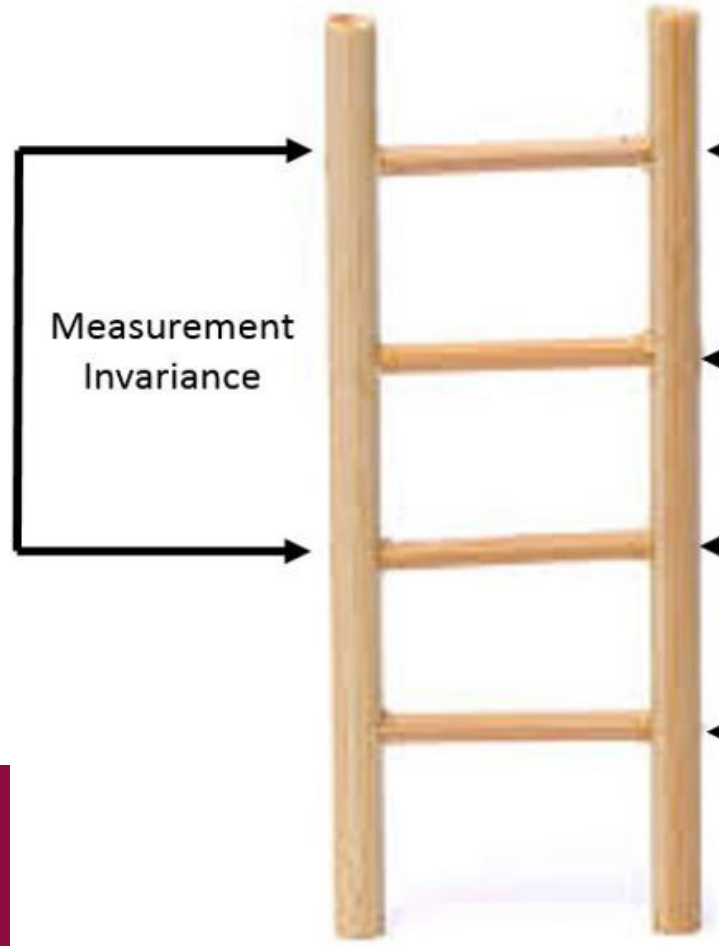
- Measurement Invariance Testing
  - Assessing the psychometric equivalence of a measure or construct across groups or across time.
  - Seeking to demonstrate that a construct has the same meaning across groups or across repeated measures.
  - Can be tested in an item-response theory or a structural equation modeling (SEM) framework.
    - Currently, SEM is the more widely used approach

Table 2

Quick Reference: MG-CFA Interpretive Guidelines

Fit Index	Interpretative Guideline
Comparative fit index (*CFI*)	< Δ 0.010
Root mean square error of approximation (*RMSEA*)	< Δ 0.015
Standardized root mean square residual (*SRMR*)	< Δ 0.030 for metric invariance and < Δ 0.015 for scalar invariance
McDonald's Noncentrality Index (*McNCI*)	< Δ 0.020

Measurement Invariance Ladder



Level of Invariance	Model Constrains	Implications for Use
Strict Invariance	Factor structure Factor loadings Item intercepts Unexplained item variance	Differences in scores between comparison groups is entirely (or at least almost entirely) due to differences in the latent variable.
Scalar Invariance	Factor structure Factor loadings Item intercepts	Differences in scores between groups reflect true variations of the underlying construct. Allows one to compare latent factor means, variances, and co-variances.
Metric Invariance	Factor structure Factor loadings	The strength of the associations between test items and the corresponding factors are statistically similar to a reasonable degree.
Configural Invariance	Factor structure	There is a common model structure between groups.

A Brief Note on Test Score Reliability

- Consistency or stability of test scores.
  - To what extent would a test taker score the same if they took the test over, and over, and over again?
- Contrast with *validity* evidence of test scores.
- Several reliability estimation methods are available.
  - Internal consistency reliability is a popular method
    - Cronbach's Coefficient Alpha (α) vs. McDonald's Coefficient Omega (ω)
- For more on reliability: https://doi.org/10.1080/07481756.2023.2283637

References

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. https://www.apa.org/science/programs/testing/standards

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, C. C., Lau, J. M., Richardson, G. B., & Dai, C.-L. (2020). Measurement invariance testing in counseling. *Journal of Professional Counseling, Practice, Theory, & Research, 47*(2), 89–104. https://doi.org/10.1080/15566382.2020.1795806

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121-149. doi:10.1177/0748175610373459

Kalkbrenner, M.T. (2024). Choosing between Cronbach's coefficient alpha, McDonald's coefficient omega, and coefficient H: Confidence intervals and the advantages and drawbacks of interpretive guidelines. *Measurement and Evaluation in Counseling and Development. 57*(2), 93-105. https://doi.org/10.1080/07481756.2023.2283637

Kalkbrenner, M.T. (2024). Choosing between Cronbach's coefficient alpha, McDonald's coefficient omega, and coefficient *H:* Confidence intervals and the advantages and drawbacks of interpretive guidelines. *Measurement and Evaluation in Counseling and Development. 57*(2), 93-105. https://doi.org/10.1080/07481756.2023.2283637

Lenz, A. S. (2025). Making sense of test score validity in counseling assessment. *Measurement and Evaluation in Counseling and Development, 58*(1), 1-6. https://doi.org/10.1080/21501378.2022.2029411

Neukrug, E.S., & Fawcett, C.R. (2020). *Essentials of testing and assessment: A practical guide for counselors, social workers, and psychologists* (Enhanced 3rd ed.). Cengage

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*(1), 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Sink, C.A., (2016). Advanced research design and assessment [PowerPoint Slides]. Old Dominion University.

Steger, M. (2007). Structural Validity of the Life Regard Index. *Measurement and Evaluation in Counseling and Development*, *40*(2), 97–109. https://doi.org/10.1080/07481756.2007.11909808

Contact Information

Mike Kalkbrenner

Professor

Department of Counseling and Educational Psychology

College of Health, Education, and Social Transformation

New Mexico State University

Email: mkalk001@nmsu.edu

Website: https://cep.nmsu.edu/facultydirectory/dr.-michael-kalkbrenner.html

Questions or Comments