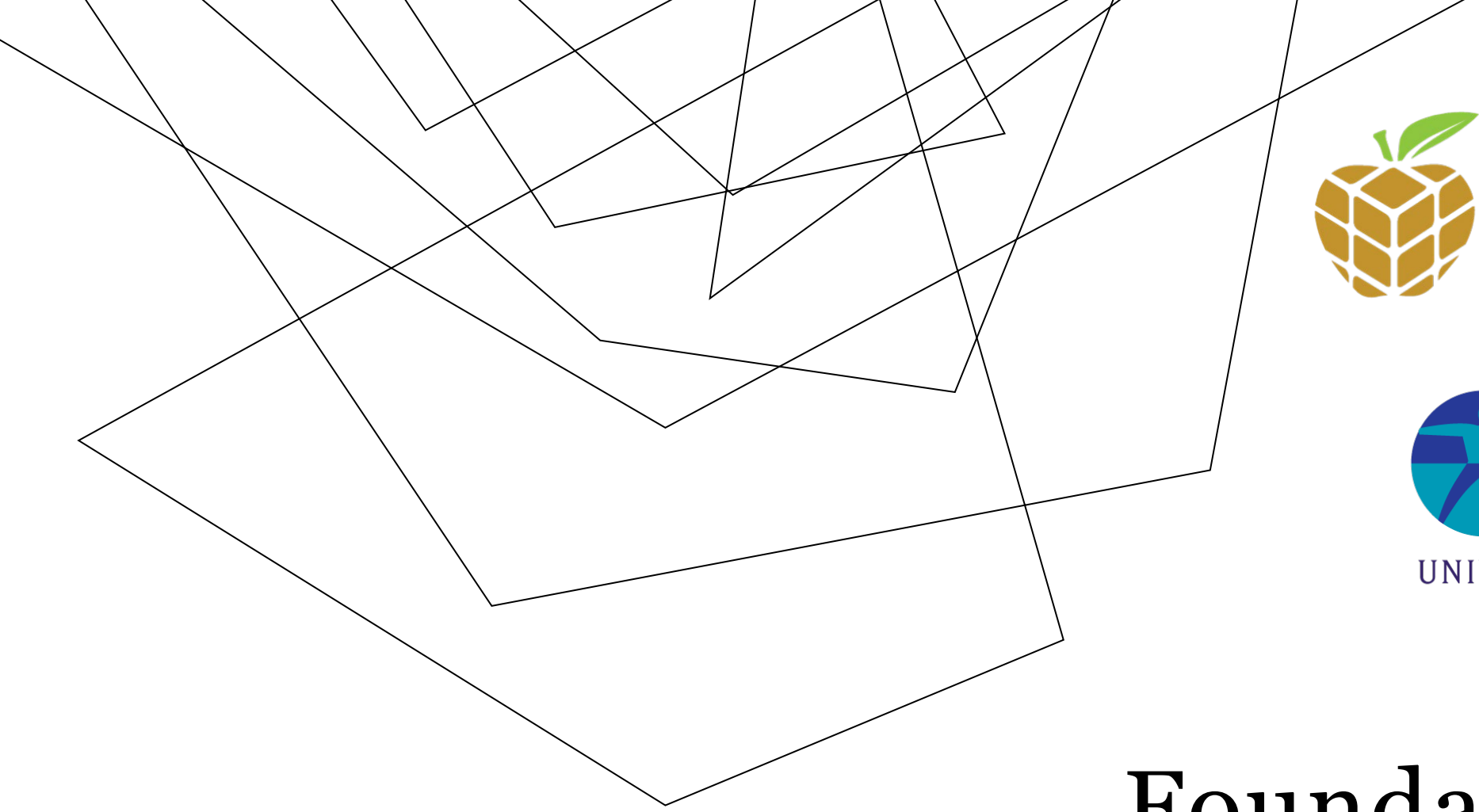


SMART
School Mental Health Assessment
Research & Training Center

HARBORVIEW
INJURY PREVENTION
& RESEARCH CENTER
UNIVERSITY of WASHINGTON

Foundations of Confirmatory Factor Analysis

KEITH HULLENAAR, PHD
QUANTITATIVE METHODS ENJOYER
UNIVERSITY OF WASHINGTON
SMART CENTER & HIPRC
PSYCHIATRY AND BEHAVIORAL SCIENCE

SMART
School Mental Health Assessment Research & Training Center

HARBORVIEW INJURY PREVENTION & RESEARCH CENTER
UNIVERSITY of WASHINGTON

Land Acknowledgement

We live and work on the **unceded** ancestral lands of the **Coast Salish people**, the land which touches the shared waters of all tribes and bands within the **Duwamish**, **Puyallup**, **Suquamish**, **Tulalip** and **Muckleshoot** nations, and pay our respects to elders past and present.

We work to create inclusive and respectful partnerships that **honor Indigenous cultures, histories, identities, and sociopolitical realities**, that dismantle ongoing legacies of settler colonialism, and that recognize the hundreds of Indigenous Nations who continue to resist, live, and uphold their sacred relations across their lands.

Welcome to SMARTSTATS

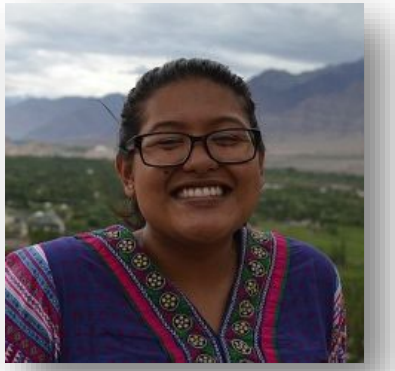
Our **mission** is to make quantitative methodologies freely accessible to **all who want to learn**.

SMART STATS

Keith Hullenaar, PhD
SMART Center, HIPRC
Assistant Professor

Bethlehem Kebede, BS
SMART Center
Research Analyst

Casey Ehde, BA
SMART Center
Research Coordinator



Mahima Joshi, MPH
SMART Center
Research Scientist I

What you should leave with

Calibrated by tier; everyone should reach all four

Vocabulary	Loadings, residual variances, identification, factor covariances. Read and interpret CFA results.
Intuition	CFA versus EFA. Negotiating fit. Why a respecification needs a story, not an MI.
Workflow	Specify, fit, inspect local fit, decide on respecification, validate, report.
Judgment	Looking at fit statistics holistically. Introduce alternative model specifications for different data.

From EFA to CFA: a change in epistemic stance

Two methods, two questions, one workflow

Exploratory Factor Analysis

"What structure do these items suggest?"

- All items load on all factors
- Rotation chooses simple structure
- Output is a hypothesis, not a test
- Used early in scale development

Confirmatory Factor Analysis

"Does this hypothesized structure fit the data?"

- Number of factors specified in advance
- Pattern of loadings specified in advance
- Cross-loadings typically fixed to zero
- Fit and parameter estimates are testable claims

Our example: the IPIP Big Five

Why this dataset is well-suited to teach CFA

n = 19,719 from openpsychometrics.org (public domain)

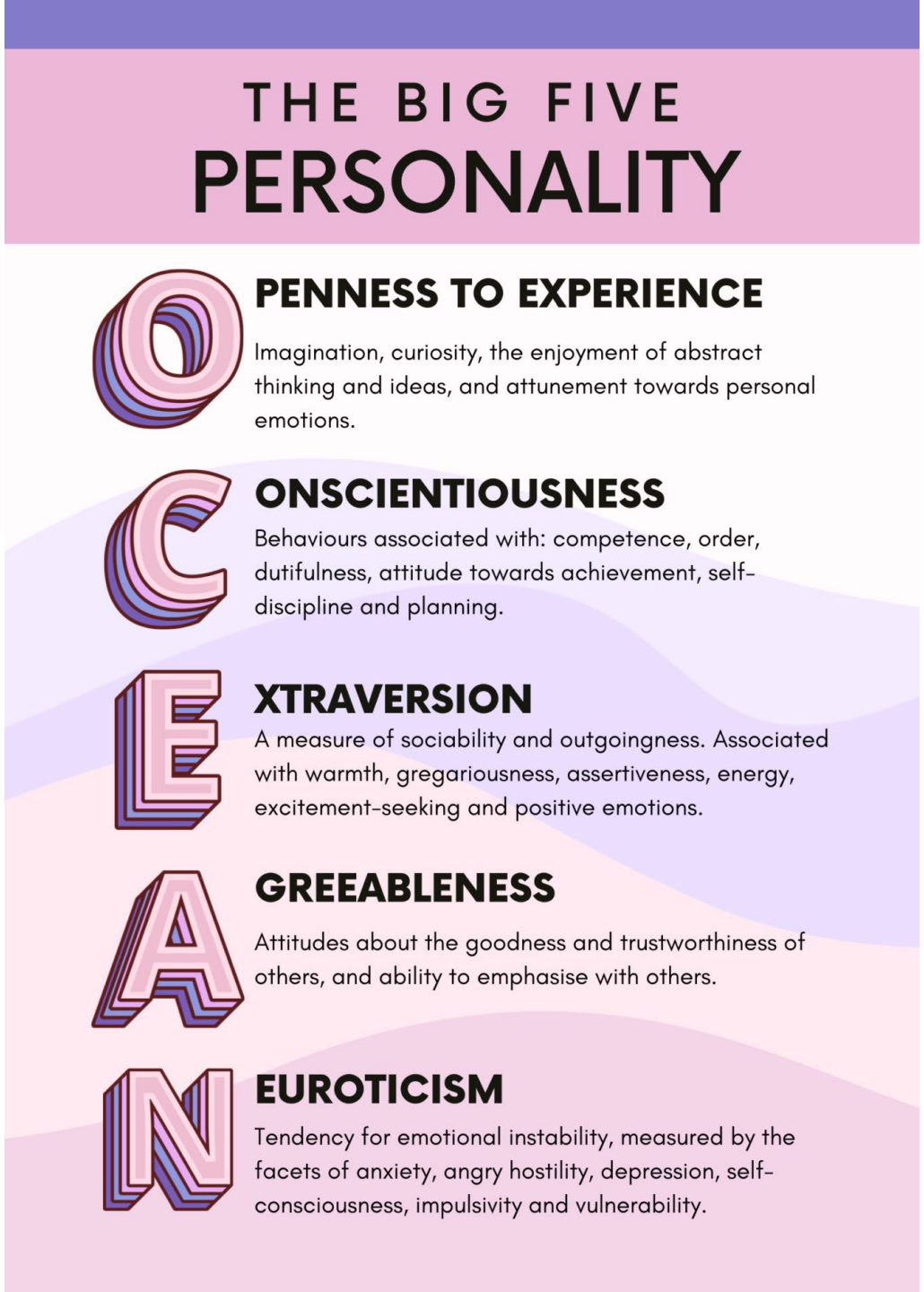
50 items, 5 factors,10 per factor:

Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), Neuroticism (N),

5-point Likert (1 = strongly disagree, 5 = strongly agree)

Why this dataset is the right teaching case for CFA:

- It's popular
- Hypothesized structure is theoretically grounded and pre-existing
- Large n surfaces the chi-square sensitivity issue honestly
- Sample size supports split-half validation



The CFA model: theory expressed as testable structure

What we mean by 'specifying a model'

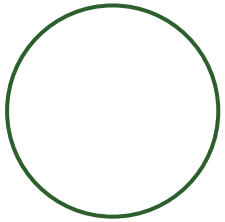
A CFA model is a set of restrictions on the relationships among observed and latent variables.

- How many factors are there
- Which items load on which factors
- Which items are allowed to share residual variance
- Whether factors are correlated

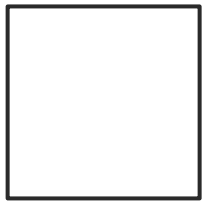
Every restriction is a claim about how the data were generated. The model fits when the data are consistent with those claims.

Path diagram conventions

Read the picture before you read the formula

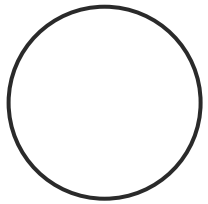


Latent variable (factor)

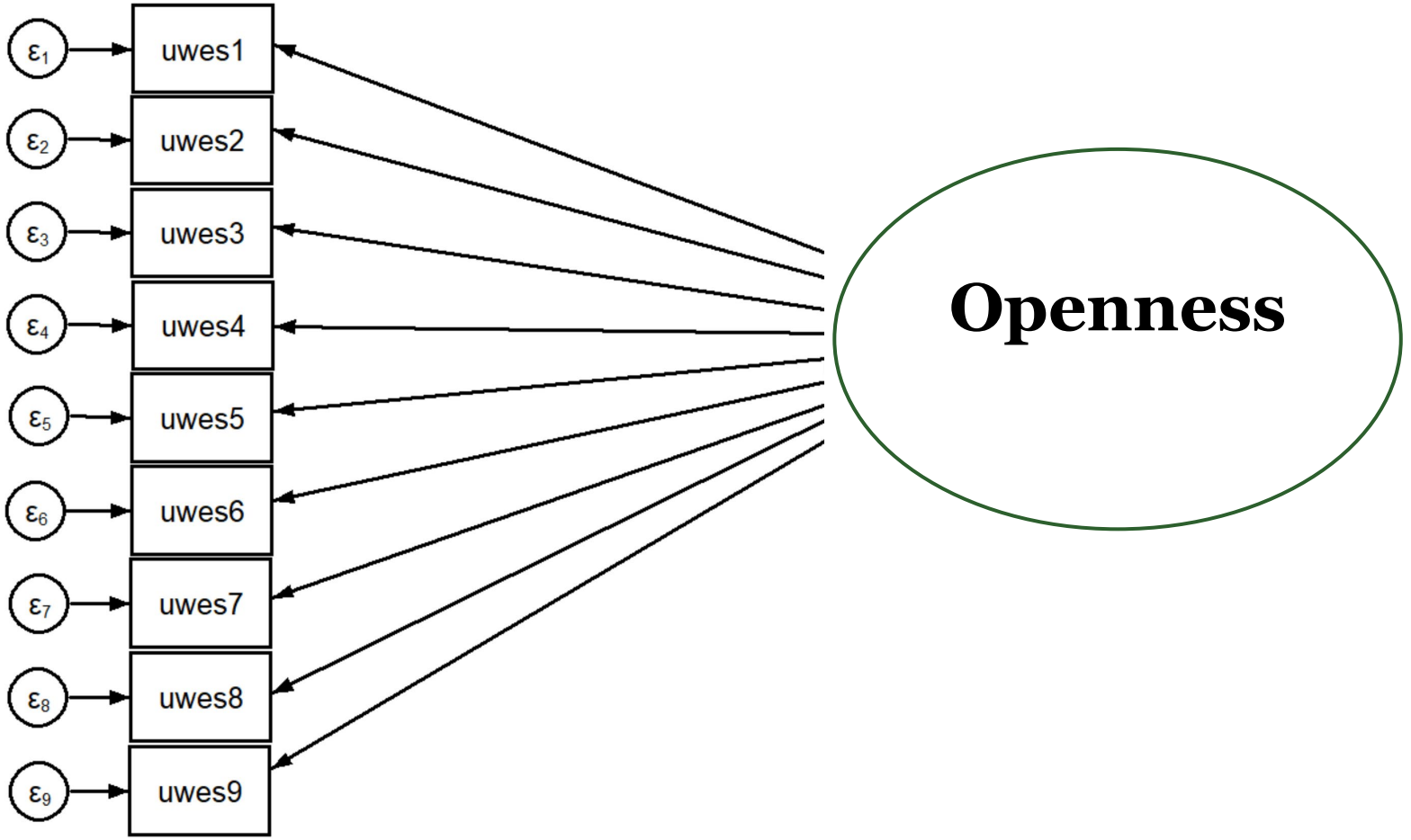


Observed indicator

Path / loading (regression)



Error



Arrows go from factor to indicators: the factor is the "common cause."

The measurement equation

One item, one latent predictor

$$y_i = \tau_i + \lambda_i \eta + \varepsilon_i$$

y_i — score on item (observed)

τ_i — intercept of item (mean of item in the absence of factor)

λ_i — factor loading of item i on factor η (how item changes when factor changes)

η — latent factor (unobserved thing we think explains the item)

ε_i — residual: variance in item not explained by the factor

Like multiple regression, but the predictor is latent.

System of equations for multiple items and one latent factor

Three items, one latent predictor

$$y_1 = \tau_1 + \lambda_0 \eta + \varepsilon_1$$
$$y_2 = \tau_2 + \lambda_0 \eta + \varepsilon_2$$
$$y_3 = \tau_3 + \lambda_0 \eta + \varepsilon_3$$

What CFA is actually doing (don’t worry about notation, right now)

Reproducing the observed covariance matrix from a few parameters

$$\Sigma(\theta) = \Lambda \, \Psi \, \Lambda' + \Theta_{\varepsilon}$$

Σ(θ): the covariance matrix the model implies given its parameters

Λ: the matrix of factor loadings

Ψ: factor variances and covariances

Θ_{ε}: residual variances (and any specified residual covariances)

Estimation finds the parameters that make Σ(θ) most resemble the observed S.

The big idea: CFA is beauty in parsimony

Why CFA is different from saturated regression

In our Big 5 data, **the 50 survey items have a 50 × 50 covariance matrix with 1,275 unique elements (variances + covariances)**

The Big 5 model suggests that this **1,275 pattern can be explained by 5 underlying factors represented by the 50 items**

If the model is right, **you can reconstruct 1,275 covariances from 110 parameters.:**

- 110 = 45 item loadings (5 are fixed to 1, ignore for now); 50 item residuals; 5 factor variances; 10 factor covariances

IN SHORT:

- A 5-factor CFA will try to reproduce those 1,275 variances+covariances using ~110 parameters.
- If it fits well, the model is defensible.
- If it does not fit well, either the model is wrong or the constraints are too tight.

SPECIAL NOTE IN MODEL BUILDING:

Fewer parameters are more testable, but harder to fit; more parameters are easier to fit, but less informative (and useful).

Two ways to identify a factor

You must scale the latent variable somehow

Marker method

Fix one loading per factor to 1.

- lavaan default
- Factor variance is freely estimated
- Other loadings are interpreted relative to the marker
- Use when factor scaling is anchored to a meaningful indicator

Variance standardization

Fix factor variance to 1; freely estimate all loadings.

- Set in lavaan with std.lv = TRUE
- All loadings are freely estimated
- Loadings are on a comparable metric within a factor
- Standard for measurement invariance

The two approaches give equivalent fit; only the scaling differs.

The 3-indicator rule

And why two-indicator factors are fragile

- **A factor with three or more indicators is identifiable on its own.**

- Two-indicator factors are only identifiable, but it's very difficult
- Even then, they are prone to estimation problems and Heywood cases.

- **If you can write three indicators (items), write three (or more).**

  If you cannot, document why and expect reviewer skepticism.

Estimation: maximum likelihood

What ML is doing, what it assumes

- **ML finds parameter values that make the observed data most likely under the model.**

Iterative procedure: starts with initial estimates, refines until Σ(θ) cannot get any closer to the observed covariance matrix.

Assumptions:

- Continuous indicators (we have tools to address)
- Multivariate normality (we have tools to address)
- Large N

Problems when assumptions fail:

- Standard errors and chi-square become biased under non-normality
- ML may not appropriate for ordinal indicators with few categories

When ML isn't enough: brief tour

Two alternatives every CFA user should recognize

MLR

Robust ML — continuous indicators, non-normal data

- Same parameter estimates as ML
- Sandwich-style adjusted standard errors
- Default for non-normal continuous outcomes in modern practice

WLSMV

Mean- and variance-adjusted weighted least squares — ordinal data

- Treats Likert items as ordered categorical
- Fits a polychoric correlation model
- Can’t use FIML missing data techniques
- Modern standard for ordinal CFA
- Slower; needs adequate N per category

Whichever you use, name it in the methods. Reviewers will ask.

Things that go wrong: Heywood cases & non-convergence

What lavaan tells you when something is off

Heywood cases

- Negative residual variances or standardized loadings > 1
- Usually indicate misspecification, two-indicator factors, or small N
- Do not 'fix' by constraining the parameter without also asking why
- You have to address it

Non-convergence

- Solver hits the iteration limit without stabilizing
- Common causes: poor starting values, weak factors, not enough indicators, bad model
- Sometimes solved by raising the iteration cap; usually a sign of trouble

Always read the warning text.

Evaluating model fit

We treat these indicators as snapshots of the holistic picture of fit (with our theory). Report all.

Absolute / exact	$Model\ \chi^2$ $SRMR$	**Chi-square** Tests S = Σ(θ) exactly. Almost always rejects at large N. **SRMR** Difference between implied and observed matrices
Approximate, parsimony	$RMSEA + 90\%\ CI$	Per-df discrepancy. Penalizes complexity.
Approximate, comparative	CFI, TLI	Improvement over a null (independence) baseline.
Local	$Standardized\ residuals, MIs$	Where, specifically, the model fails.

No single index settles it. Disagreement among indices is itself information.

Model chi-square : useful and dangerous

Why p < .001 is not the verdict you think it is

Tests the null that S = Σ(θ) exactly.

- At large N, even tiny deviations produce significant chi-square.
- At small N, the test may lack power to detect real misfit.
- Distribution assumptions matter; non-normality biases the test.

What it IS still good for:

- Comparing nested models ($\Delta\chi^2$ difference test)
- Sanity-checking that your output is structurally what you expected (df count)

We rarely use chi-square alone to assess fit. We use RMSEA, SRMR, TFI, CLI.

SRMR: standardized root mean square residual

An average of how far off you are, in correlation units

Average absolute discrepancy between observed and model-implied correlations.

- Range 0 to 1; smaller is better.
- Interpretable on the correlation metric: an SRMR of .05 means residual correlations average about .05.

Conventional cutoff: SRMR ≤ .08.

Why include it :

- Catches misfit the others miss — especially when models have many indicators
- Less sensitive to model size than RMSEA
- Behaves poorly under categorical indicator estimators (WLSMV)

Comparative fit: CFI and TLI

Improvement over a model that says nothing

Both compare your model against a baseline that fixes all covariances to zero (the 'null' model).

- CFI: Comparative Fit Index. Range 0 to 1. Higher is better.
- TLI: Tucker–Lewis Index. Similar interpretation, penalizes complexity, can exceed 1.

Conventional cutoffs (Hu & Bentler, 1999): CFI ≥ .95, TLI ≥ .95 for 'good' fit.

Watch out:

- Both depend on a sensible null model. With many highly correlated items, the null is easy to beat.
- When CFI and TLI disagree, the model is borderline. Don't pick the friendlier number. Report both.

RMSEA and the close-fit test

Per-df misfit, with a confidence interval and a hypothesis test

RMSEA estimates discrepancy per degree of freedom in the population.

- Comes with a 90% confidence interval. Always report it.
- Comes with the close-fit test: H_0 that population RMSEA ≤ .05.
- If the close-fit p-value is non-significant, you cannot reject 'reasonable fit.'

Conventional thresholds:

≤ .05 close fit; ≤ .08 reasonable; ≤ .10 marginal; > .10 poor

Caveats:

- Behaves poorly with low df and small N
- Like all fixed cutoffs, depends on the model conditions in the simulations that produced them

The cutoff debate you should know about

Hu & Bentler (1999) was a starting point, not the verdict

- **The standard cutoffs (CFI ≥ .95, RMSEA ≤ .06, SRMR ≤ .08) come from Monte Carlo simulations on a specific model class.**

- Marsh, Hau, & Wen (2004): cutoffs depend on model size, factor reliability, and other features.
- McNeish & Wolf (2023, Psych Methods): fixed cutoffs misbehave for typical psychological CFAs.
- Groskurth et al. (2024): independent simulation reaches the same conclusion.

- **What this means for your workshop CFA:**
- Report fit indices with their conventional cutoffs.
- Do not treat any single index as decisive.
- Flag local misfit honestly; cite the cutoff debate when reviewers push back.

Local fit: where the model actually fails

Global fit averages over many things; locally, you see them

Standardized residual covariances

- Each cell: how badly the model reproduces that pair of items
- Values |z| > ~2.5 flag pairs the model is misfitting
- Cluster patterns suggest cross-loadings or method effects

Modification indices (MIs)

- Expected drop in chi-square if a fixed parameter were freed
- Large MI = the data 'want' that parameter estimated
- Available for cross-loadings, residual covariances, and structural paths

The respecification trap

Why MI-driven model search is capitalization on chance

If you free every parameter the MI suggests, your model will fit — on this sample.

- It will not generalize.
- It is no longer a confirmatory test; it is exploratory regression dressed in CFA notation.

Three criteria for a defensible respecification:

- Substantive rationale independent of the MI (theory, item content, known method effect)
- You would have predicted this parameter before seeing the MI
- You will report and justify it transparently in the paper

If thy answer is no, thou do not free the parameter.

Section 3

Applied example in R

Specify → fit → evaluate → decide → validate → report

Resources

Five readings to take home

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research (2nd ed.). Guilford Press.*

Great text. Chapters 3 and 4 cover everything in this session at depth.

Roos, J. M., & Bauldry, S. (2022). *Confirmatory Factor Analysis. SAGE Publishing.*

Modern CFA textbook

Hu, L., & Bentler, P. M. (1999). *Cutoff criteria for fit indexes… SEM, 6(1), 1–55.*

The source of the conventional cutoffs you will see in every published CFA.

McNeish, D., & Wolf, M. G. (2023). *Dynamic fit index cutoffs… Psychological Methods, 28(1), 61–88.*

Why fixed cutoffs misbehave for typical psychological CFAs, and what to do instead.

Nye, C. D. (2023). *Reviewer resources: Confirmatory factor analysis. ORM, 26(4), 608–628.*

Concise reviewer-and-author checklist. Useful for your own papers and for reviewing others'.

Questions

Discussion welcome on anything in the slides, the script, or your own data.