NEBRASKA ACADEMY FOR METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Methodology Applications Series

2022-2023

University of Nebraska Lincoln

NEBRASKA ACADEMY FOR
METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Treatment Fidelity from a Measurement Perspective

James A. Bovaird, PhD

University of Nebraska-Lincoln

UNIVERSITY OF Nebraska Lincoln

Acknowledgement:

Jonah Garbin

Jay Jeffries

Sunny Lee

Amelia Miramonti

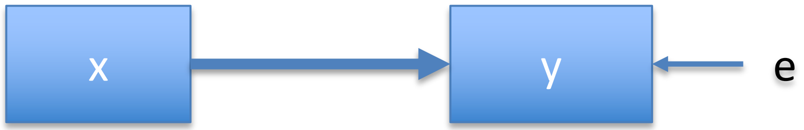
Introduction

- Treatment fidelity is a multifaceted concept that can be very difficult to measure well. This presentation will draw parallels between treatment fidelity and general psychological and educational measurement, focusing on the concepts of construct explication, reliability, validity, and the potential benefits of utilizing treatment fidelity information – or the consequences of failing to do so.

- Acknowledgement:

  Jonah Garbin
  Jay Jeffries
  Sunny Lee
  Amelia Miramonti

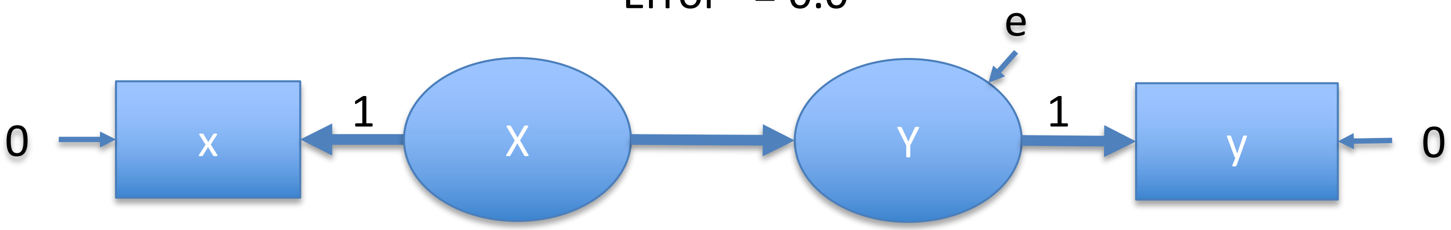
Basic Measurement Assumptions in "Traditional" Statistics



"Standard" regression, or any general linear model, assumes that the modeled variables – predictor or outcome – are measured ***without*** error.

Reliability = 1.0

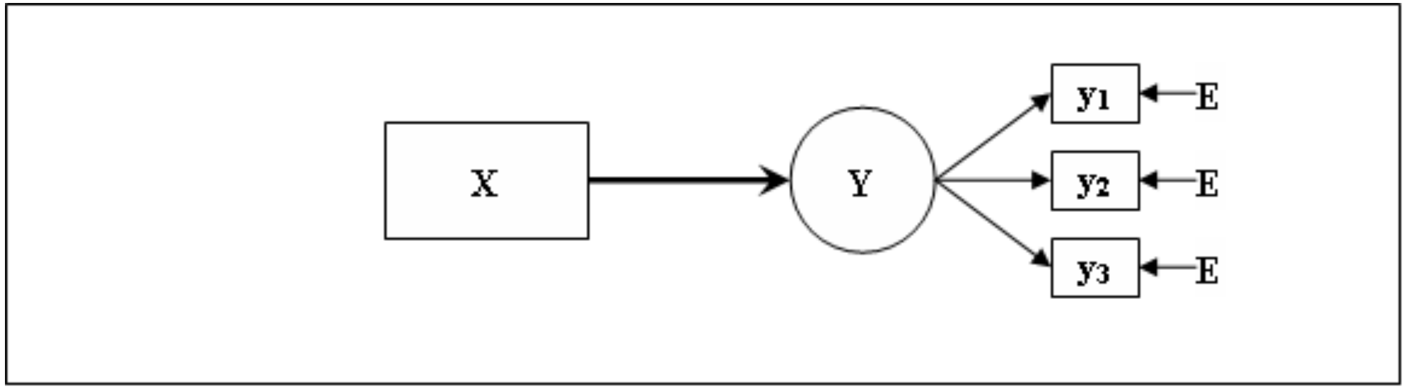
"Error" = 0.0



We ***either:***

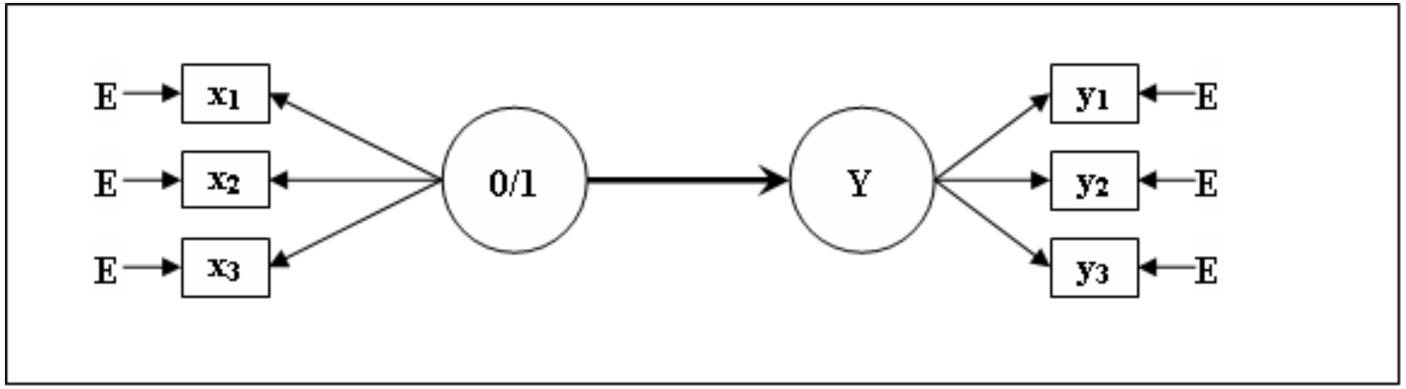
1) Work to obtain scores that ARE either perfectly reliable OR above a pre-specified threshold (i.e., *Cronbach's Coefficient* α > .75) – and proceed as "usual".
2) Stop making assumptions about the reliability of scores and correct the model for the degree of unreliability.

When we have Un-***Reliability*** in our Outcomes



MAP ACADEMY

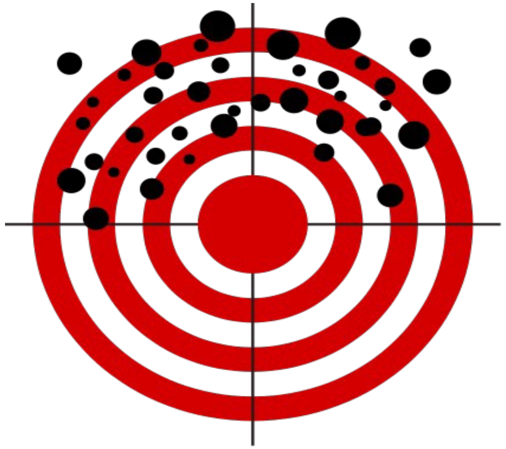
When We have Un-***Fidelity*** in Our Interventions



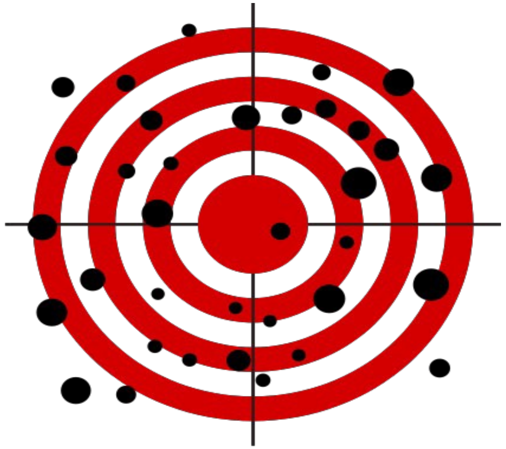
MAP ACADEMY

What is Reliability?

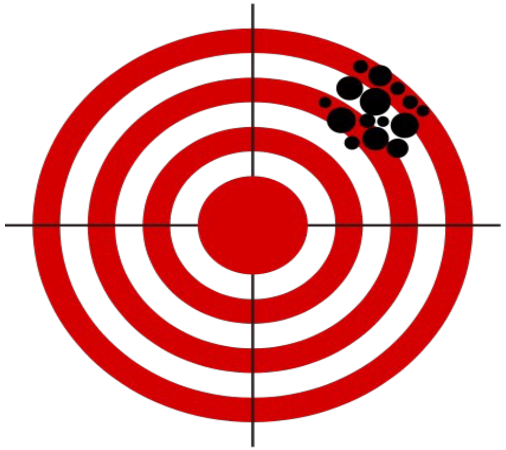
Reliability vs. Validity



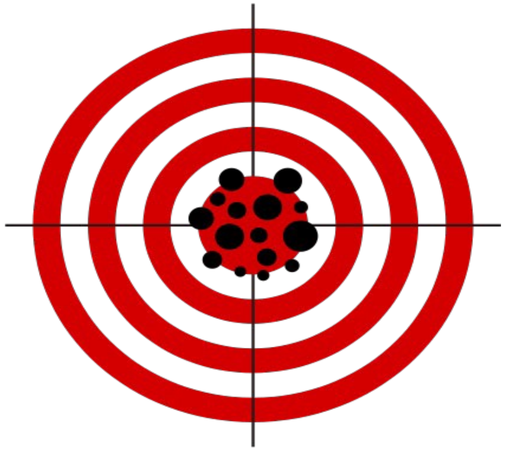
Unreliable & Invalid



Unreliable, But Valid



Reliable, Not Valid



Both Reliable & Valid

- Reliability is defined as ***consistency***
- Validity is typically presented as ***accuracy***
  - “The degree to which a test measures the construct it purports to measure”
- Test scores can be reliable but not valid

MAP ACADEMY

Attributes of Individual Items

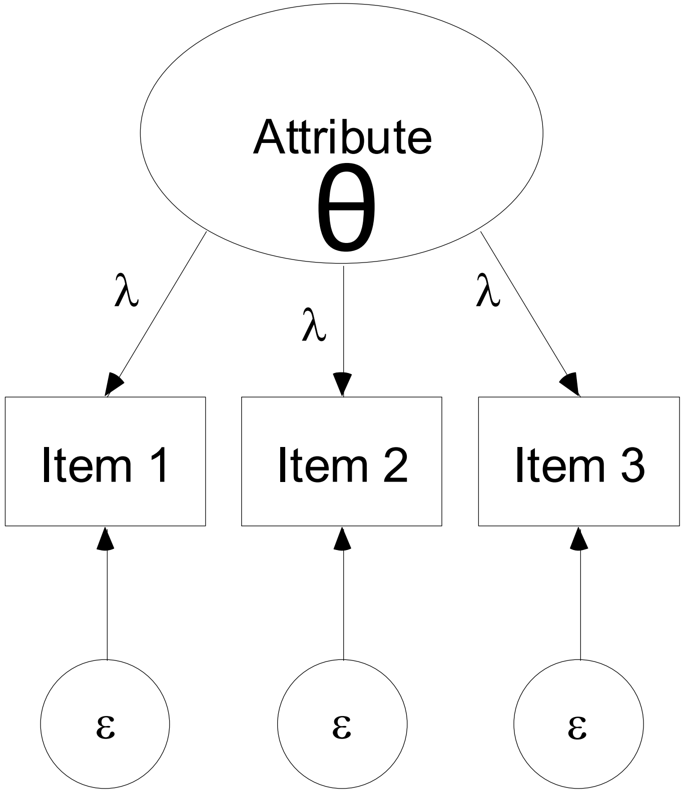
- Data from individual items …
  - Correlate poorly with the attribute of interest
  - Are unreliable
  - To some degree measure attributes other than the one of interest (i.e., contain unique variance)
  - Differ in strength or difficulty (or “quality”)
  - Categorize individuals into a small number of groups, and thus don’t provide fine discrimination

Multiple Item Measures

- The complexity (and latent nature) of the constructs studied in psychology and education generally requires the use of ***multiple*** item measures
- Individual items cannot accurately represent the construct, nor do they result in reliable data
- The number of items required depends on the complexity of the construct

A Measurement Model

- Individual items are influenced by the amount of the attribute (e.g., depression; denoted θ) one possesses
    - *“Reliable” variance*
- Items are also influenced by random error and systematic sources unrelated to the attribute (denoted ε)
    - *“Unreliable” variance*

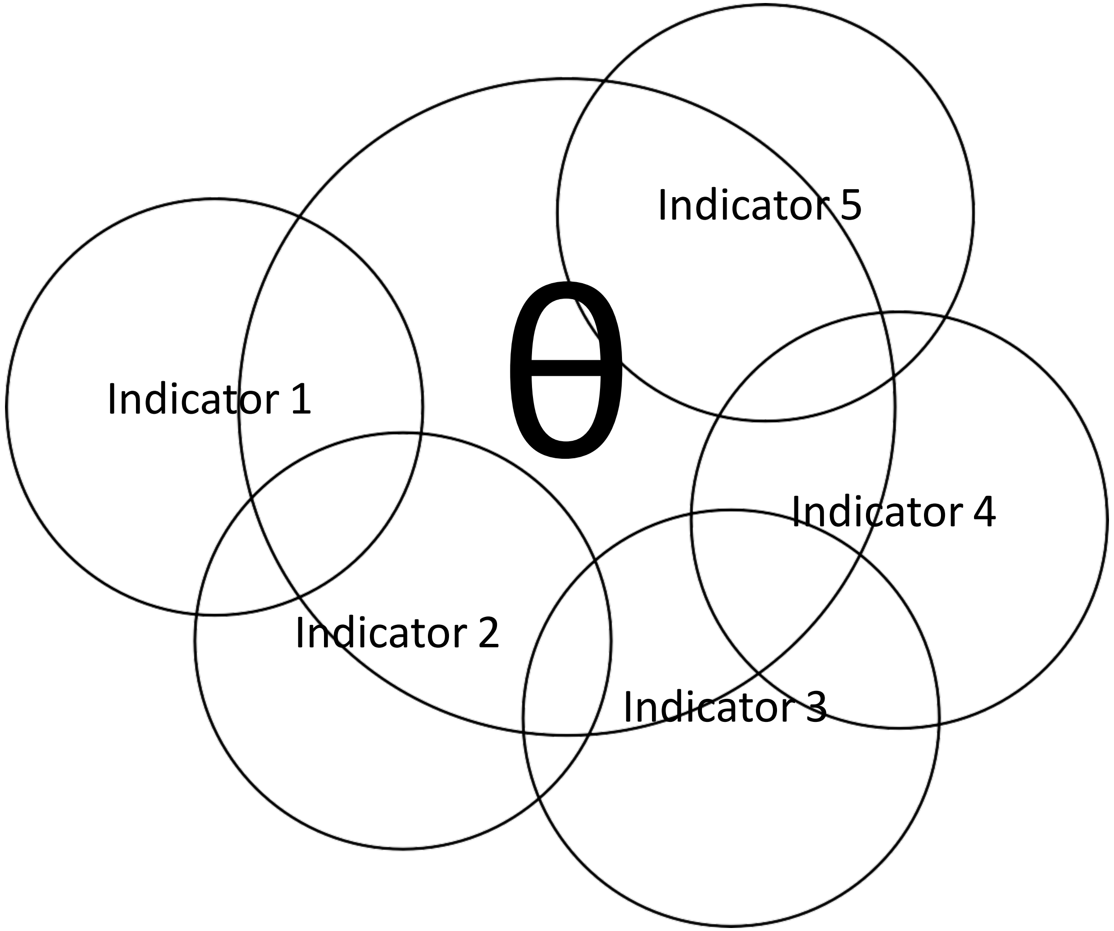


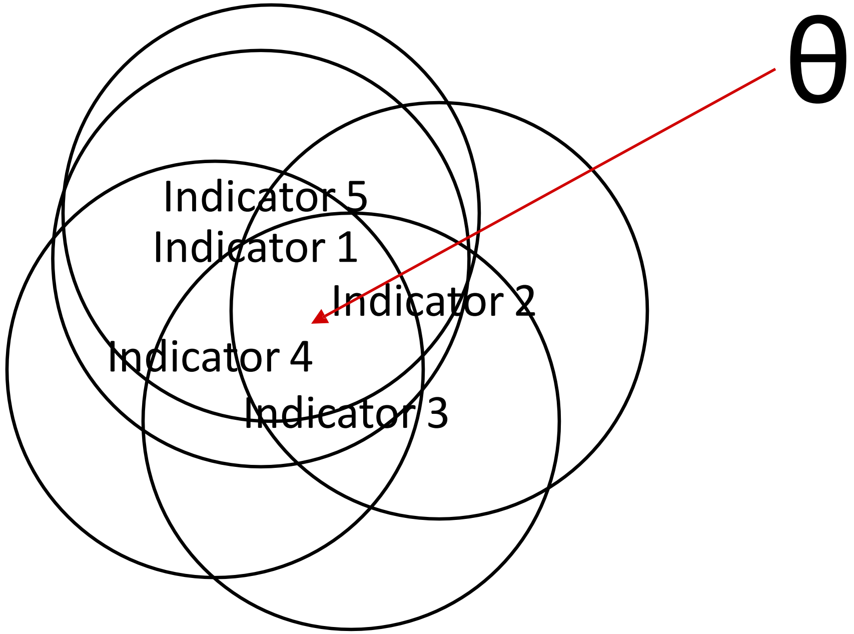
MAP ACADEMY

Reliability of Multi-Item Scores

- Of Composite Scores…
  - When there are multiple scores available for an individual one can calculate composite scores (e.g., assigning grades in class).
  - There are different approaches to calculating the reliability of composite scores, but the important issue is that the ***reliability of a composite score is generally greater than the reliability of the individual scores***.
- Of Difference Scores
  - Difference scores are calculated when comparing performance on two tests.
  - ***The reliability of the difference between two test scores is generally lower than the reliabilities of the two tests***.
  - Why? Both tests have unique error variance that they contribute to the difference score.

Conceptually… vs. In Practice…





MAP ACADEMY

Mapping Out the Construct

Content & Construct Validity

- Content Validity
  - "Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose." (Haynes, Richard, & Kubany, 1995, p. 238)

- Construct Validity
  - The term construct refers to some attribute we seek to measure
  - In social sciences, constructs (e.g., IQ) are not directly observable, but are considered latent traits
  - Construct validity attempts to answer the question, ***"What do scores on this test mean?"***

Defining the Construct

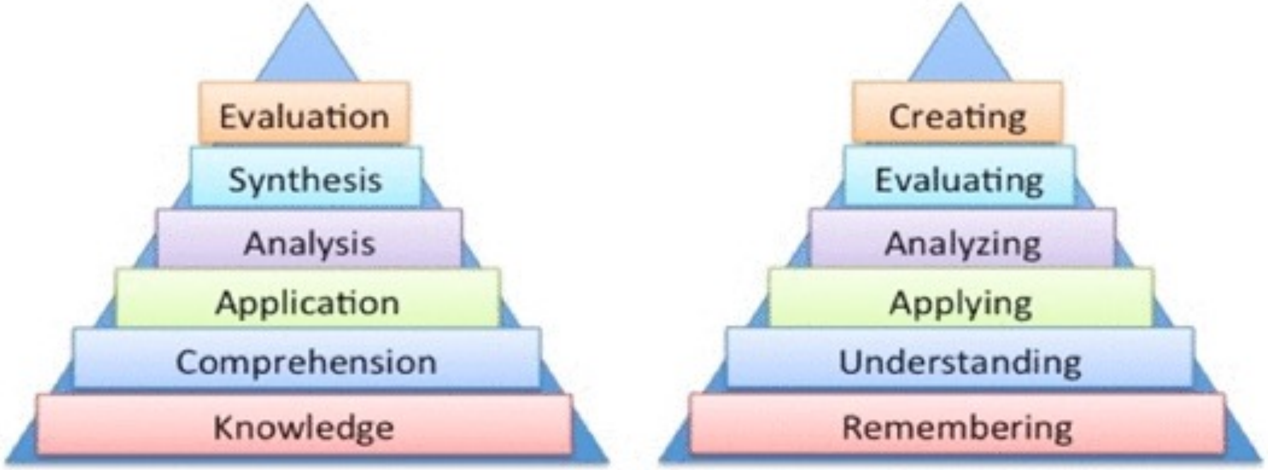
- The first step in the process of building any assessment is to ensure that you fully understand the target construct(s)
  - Definitions of what you are measuring including any theorized subdomains
- The next step is to decide the scale on which you want to locate respondents
  - Continuous? Categorical? How many categories? Norm/criterion-referenced?
  - Provide clear examples of what a respondent at each level can/cannot be expected to do (standard setting)
  - What claims do you want to make about students (content/scale)? What evidence will you need to collect in order to make those claims (tasks/items)?
- There are several popular frameworks for navigating this process:
  - Construct Mapping (Wilson, 2004)
  - Evidence-Centered Design (Mislevy, Almond, & Lukas, 2003)

Defining Test Content

- For cognitive measures, content validity follows from the test plan (i.e., table of specifications)
- A ***table of specifications*** is a blueprint that defines the cognitive processes and content covered on the exam
- The blueprint should outline the content to be tested, the types of items used, number of items devoted to each topic, etc.

Bloom’s Taxonomy

- Bloom established a hierarchy that attempts to divide cognitive tasks into categories ranging from simple to complex
- Each successive level requires the presence of lower-order skills



Bloom’s Taxonomy

21st Century Bloom’s Taxonomy

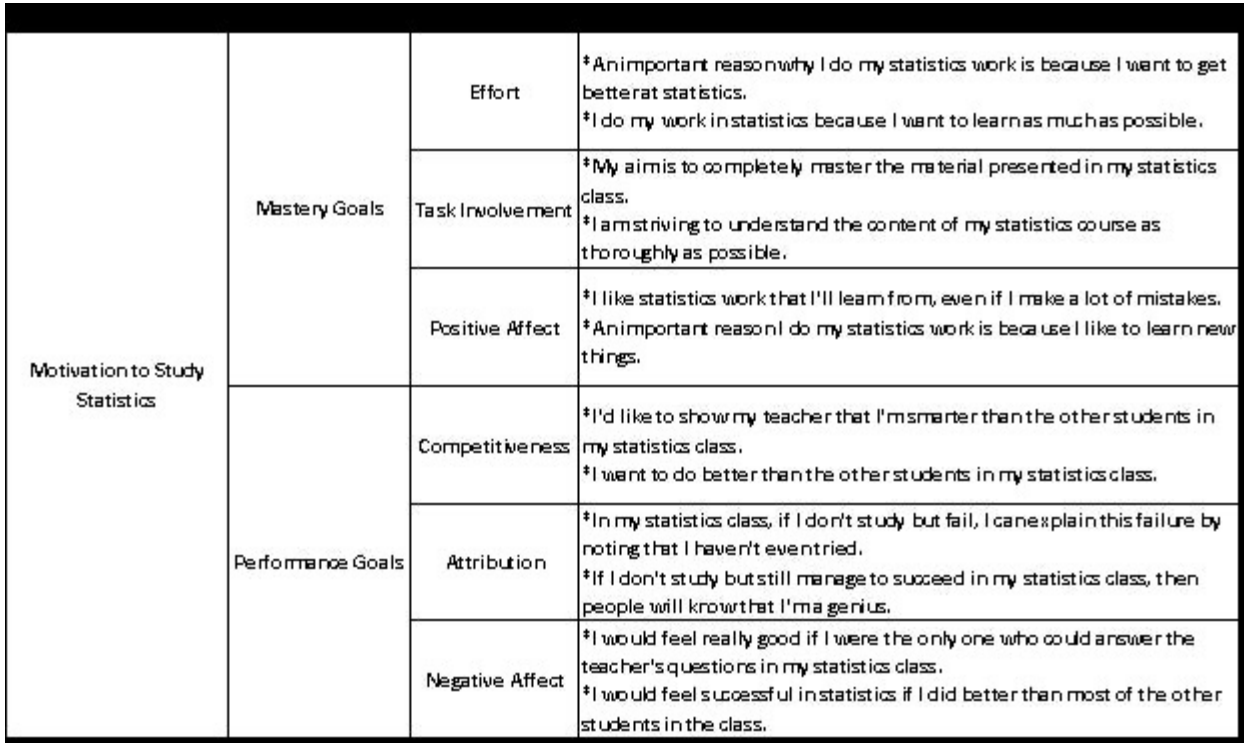
Source: Anderson & Krathwohl (2001)

Example Table of Specifications (or Concept Map)

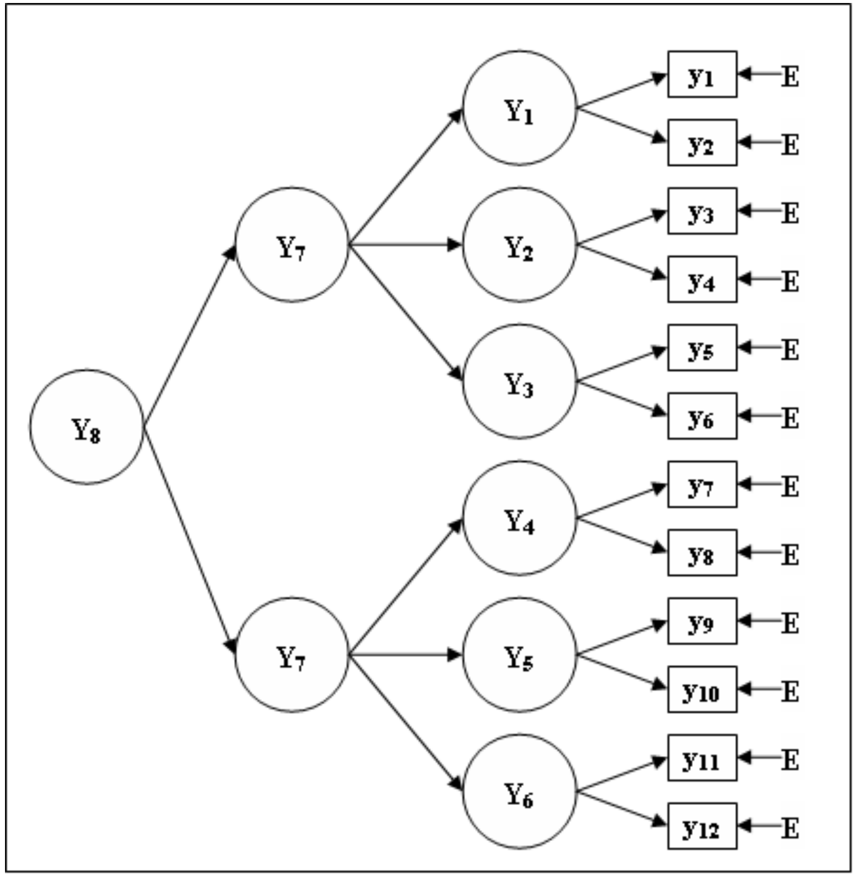
Content Areas	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Total
Scales of Measurement	**6.7%**	**6.7%**		**6.7%**			**20%**
Measures of Central Tendency	**10%**	**10%**					**20%**
Measures of Variability	**10%**	**10%**	**10%**				**30%**
Correlation & Regression	**6.7%**	**10%**		**6.7%**	**6.7%**		**30%**

Source: Reynolds, Livingston, & Willson (2009)

Concept Map → Measurement Model



Motivation to Study Statistics	Mastery Goals	Effort	‡An important reason why I do my statistics work is because I want to get better at statistics. ‡I do my work in statistics because I want to learn as much as possible.
		Task Involvement	‡My aim is to completely master the material presented in my statistics class. ‡I am striving to understand the content of my statistics course as thoroughly as possible.
		Positive Affect	‡I like statistics work that I'll learn from, even if I make a lot of mistakes. ‡An important reason I do my statistics work is because I like to learn new things.
	Performance Goals	Competitiveness	‡I'd like to show my teacher that I'm smarter than the other students in my statistics class. ‡I want to do better than the other students in my statistics class.
		Attribution	‡In my statistics class, if I don't study but fail, I can explain this failure by noting that I haven't even tried. ‡If I don't study but still manage to succeed in my statistics class, then people will know that I'm a genius.
		Negative Affect	‡I would feel really good if I were the only one who could answer the teacher's questions in my statistics class. ‡I would feel successful in statistics if I did better than most of the other students in the class.



MAP ACADEMY

Measuring Reliability

Cronbach’s Coefficient Alpha

- Cronbach’s coefficient alpha is the most common reliability estimate
  - Quantifies consistency across items within a test
  - Appropriate for multiple-item measures that are **unidimensional**
    - i.e., Measure a single common construct
  - The average of all possible Spearman-Brown-corrected split-half correlations

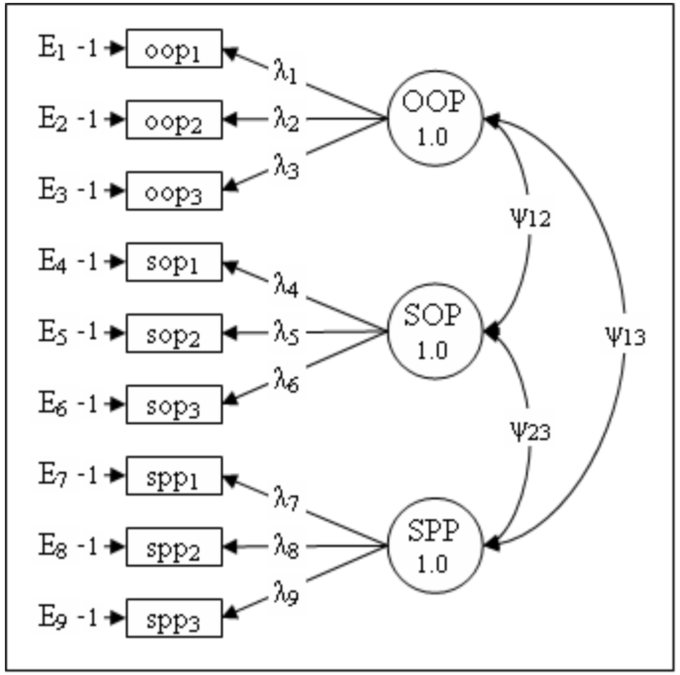
Internal Consistency Assumptions Underlying Usage of Total/Sum Scores

- Under Classical Test Theory (CTT):
  - Coefficient alpha derivations ***assume*** that all items measure the same construct (i.e., the test is ***unidimensional***)
    - Far too often, researchers use alpha without first evaluating the dimensionality of their measure.
    - Coefficient alpha is ***NOT*** a test of unidimensionality!
  - All items are ***assumed*** to be ***equally related*** to the construct (i.e., parallel measures)
    - When items are not parallel (they typically are not), alpha is a ***lower-bound*** reliability estimate (i.e., underestimates the reliability)
- Cortina (1993) covers alpha and its limitations nicely

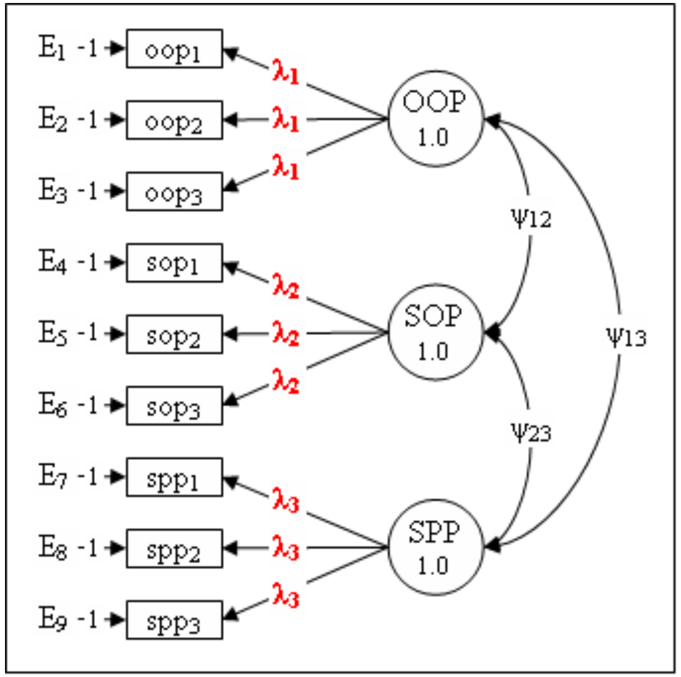
MAP ACADEMY

What are “Parallel” Items?

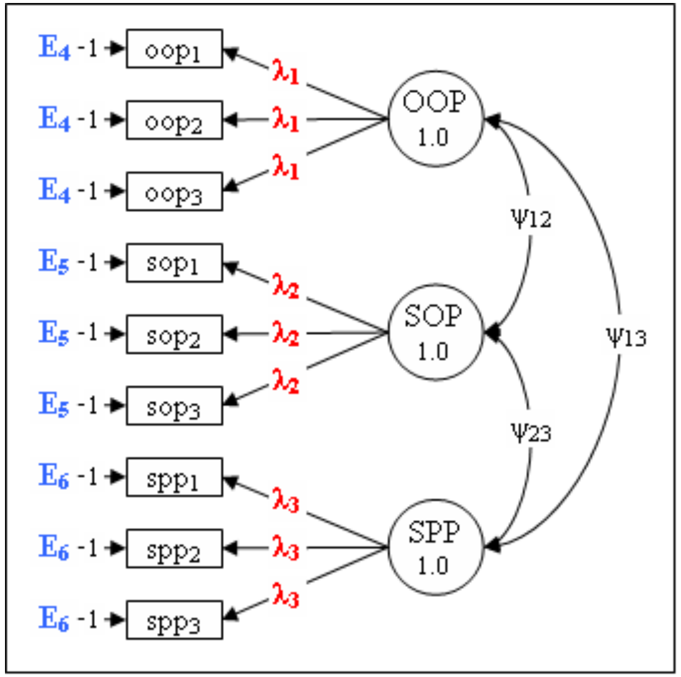
- **Congeneric**
  - Items measure the same construct but not necessarily to the same degree



- **Tau-equivalent**
  - Indicators are congeneric & have equal true score variabilities



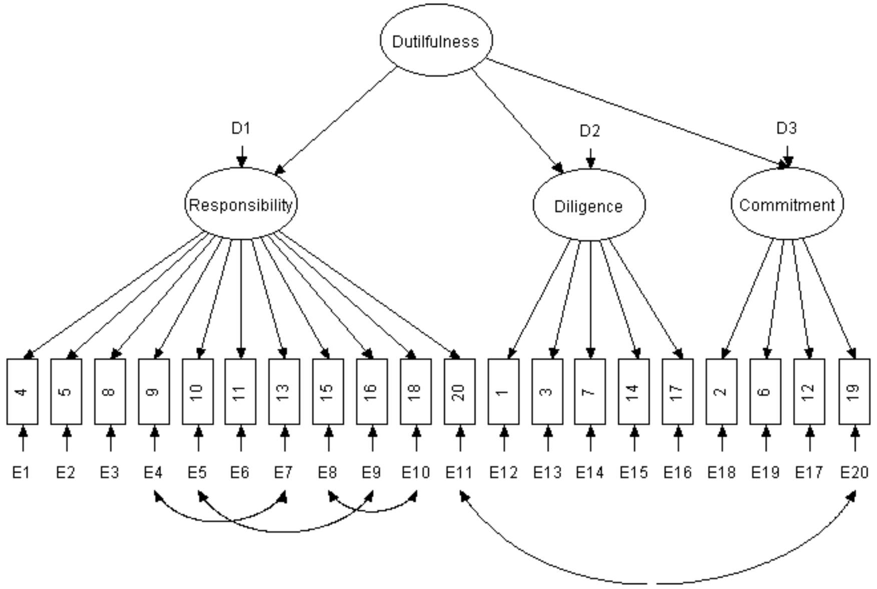
- **Parallel**
  - Equal error variances



MAP ACADEMY

Latent Variable Approaches to Scale Reliability

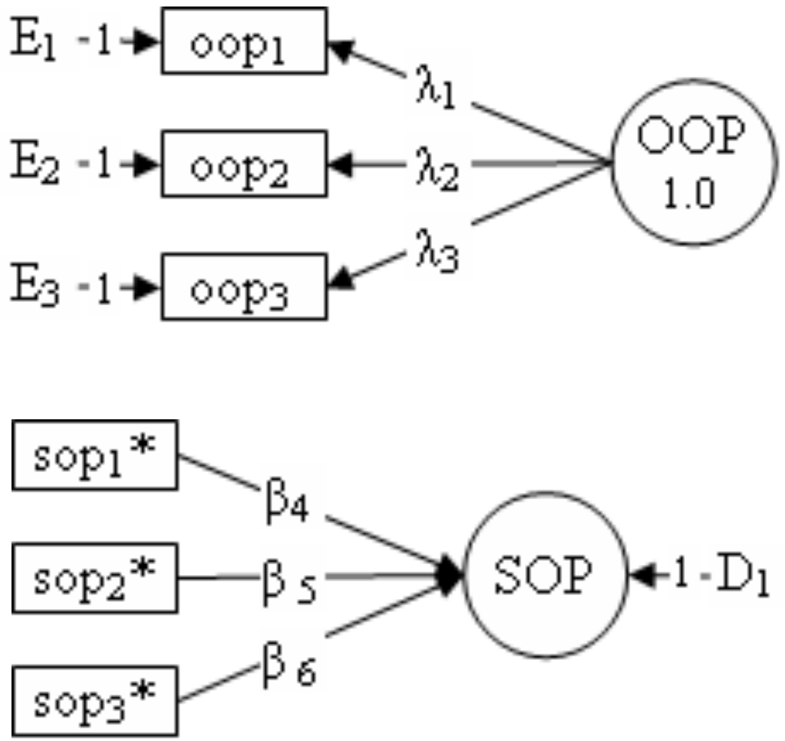
- Raykov (1997, 2001a, 2001b, 2001c, 2002) & *many* others…
- SEM method free from biases of Cronbach’s Alpha
  - Especially useful when correlated errors
  - Alpha becomes *lower bound* for scale reliability



$$\rho_Y = \frac{\left(\sum_{i=1}^{k} b_i\right)^2}{\left(\left(\sum_{i=1}^{k} b_i\right)^2 + \sum_{i=1}^{k} \theta_{ij} + 2 \sum_{1 \le i \le j \le k} \theta_{ij}\right)} = 0.971$$

Measurement Theory Plays a Role

- Role of Measurement Theory
  - interpreted as regression coefficients
  - May be standardized or unstandardized
  - Effect/reflective indicators
    - measured variable caused by the factor
  - Cause/formative indicators
    - factor caused by the measured variable
    - SES



Standards for Reliability

- If a test score is used to make important decisions that will significantly impact individuals (i.e. ***high stakes***), the reliability should be very high: > .90 or > .95
- If a test is interpreted independently but as part of a larger assessment process (e.g., personality test), most set the standard as .80 or greater.
- If a test is used only in group research or is used as a part of a composite (e.g., classroom tests), lower reliability estimates may be acceptable (e.g., .70s).

In practice….
If we meet one of these "standards", then we assume the data is "reliable-enough" and do little/nothing…

MAP ACADEMY

Correcting for Un-Reliability: Option 1

- Pearson's *r (and any measure of association or effect)* is attenuated (i.e., made smaller) by unreliable measures

The ***correction*** for attenuation gives that correlation that would be observed had the two measures been perfectly reliable

$$r'_{xy} = \frac{r_{xy}}{\sqrt{\rho_{xx'}\rho_{yy'}}}$$

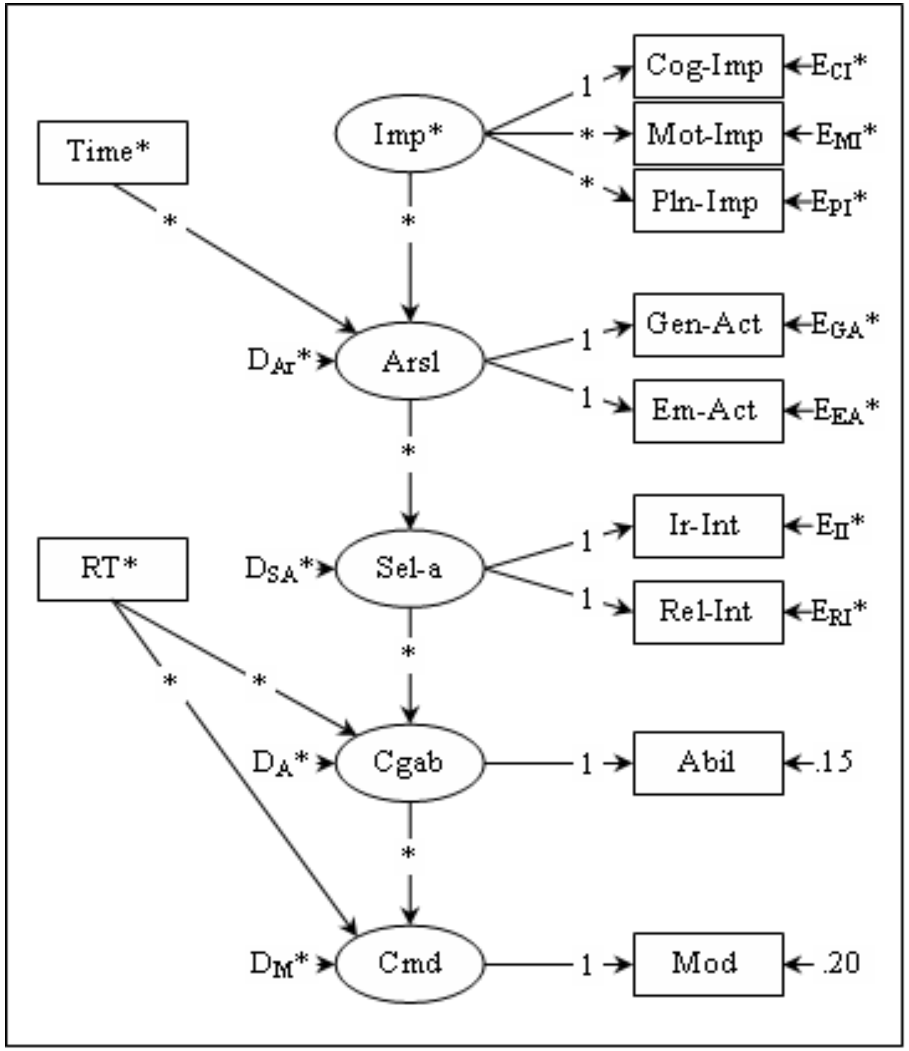
Post hoc!

Assuming perfect reliability may not be realistic. It is possible to ***correct*** the correlation using reliability values that are "more reasonable", but not perfect, say 0.90 (rather than 1.00)

$$r'_{xy} = r_{xy}\frac{\sqrt{\rho'_{xx'}\rho'_{yy'}}}{\sqrt{\rho_{xx'}\rho_{yy'}}}$$

MAP ACADEMY

Correcting for Un-Reliability: Option 2



- Instead of assuming perfect reliability, or using a post-hoc adjustment, consider ***modeling*** the un-reliability directly through a latent variable approach.

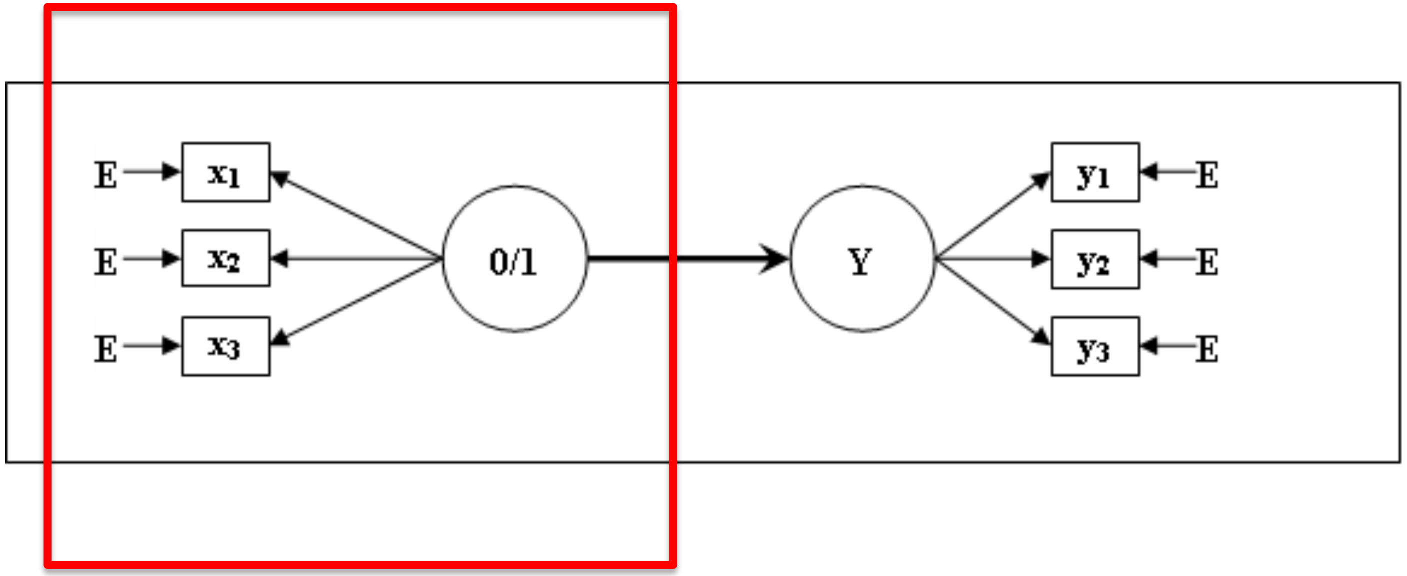
A priori & intentional!

MAP ACADEMY

What is Fidelity?

What is Fidelity?

- Fidelity is the reliability of how we implement programs, practices, interventions, etc.



MAP ACADEMY

Five Dimensions of Fidelity

Adherence: the accurate delivery of the key components of an intervention as intended.

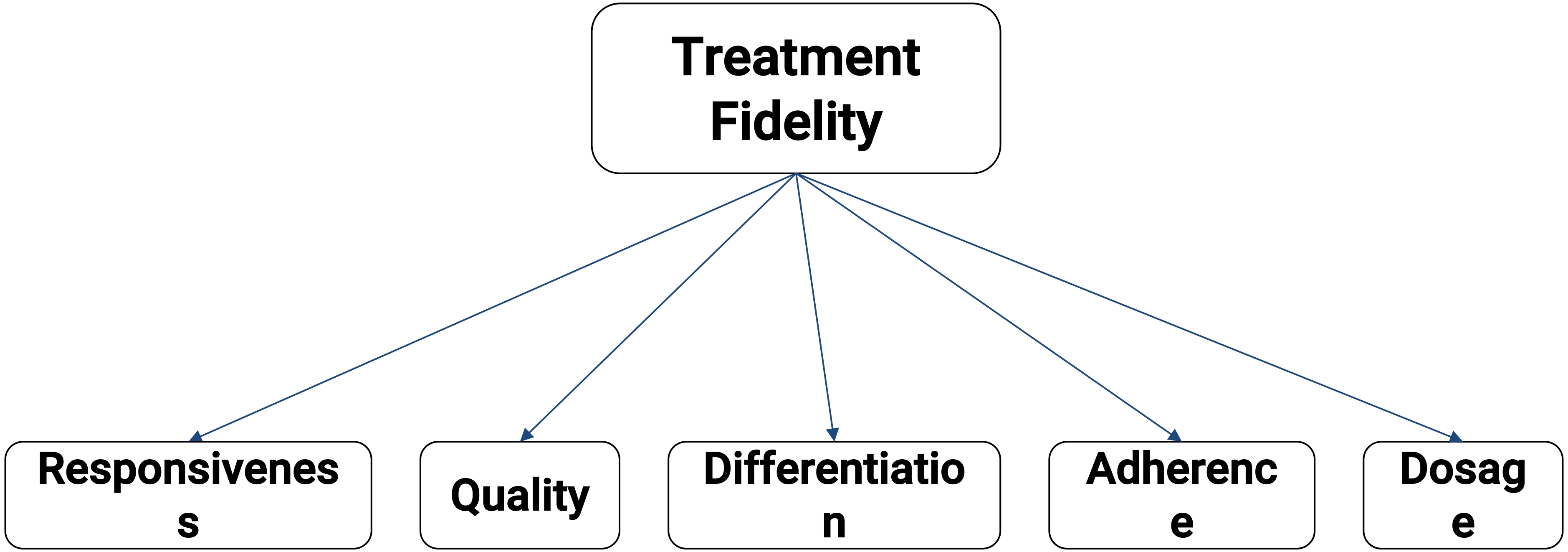
Dosage: the amount of a specific intervention delivered

Quality of intervention delivery: The way interventionists deliver the intervention using overall processes or strategies as prescribed by developers.

Participant responsiveness: The extent to which participants respond to or are engaged by the intervention and is another overall qualitative judgement.

Program differentiation: The extent to which the components and processes of the intervention being studied differ from other interventions (e.g., in a comparison of interventions study).

Dimensions of Treatment Fidelity



Treatment Fidelity

Responsivenes
s

Quality

Differentiatio
n

Adherenc
e

Dosag
e

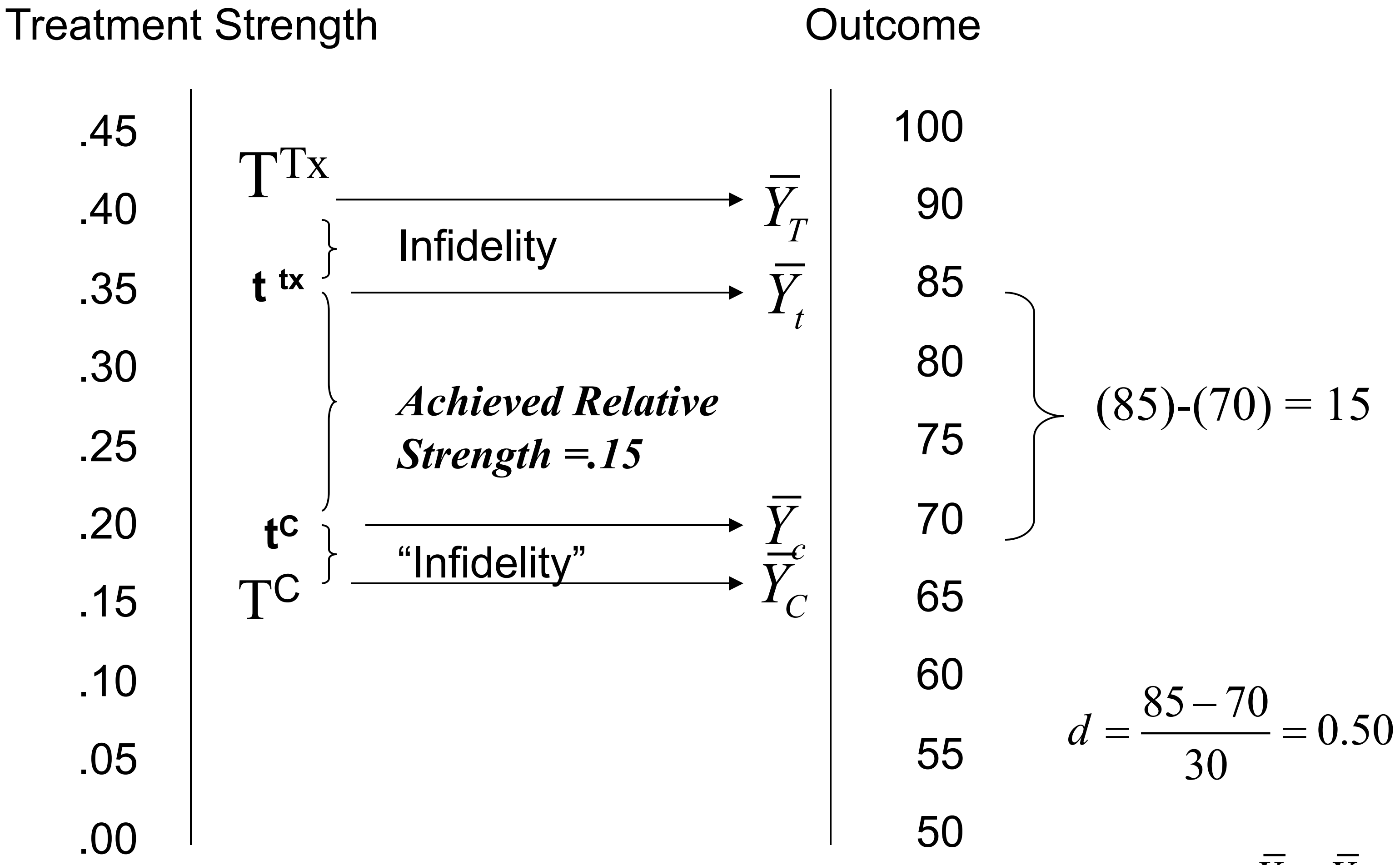
MAP ACADEMY

Defining the Fidelity Construct

Linking Intervention Fidelity Assessment to Contemporary Models of Causality

- **Rubin's Causal Model:**
  - True causal effect of X is $(Y_i^{Tx} – Y_i^{C})$
  - RCT methodology is the best approximation to this true effect
  - In RCTs, the difference between conditions, on average, is the causal effect
- **Fidelity assessment within RCTs entails examining *the difference between causal components* in the intervention and control conditions.**
- **Differencing causal conditions can be characterized as *achieved relative strength* of the contrast.**
  - **Achieved Relative Strength (ARS) = $t^{Tx} – t^{C}$**
  - **ARS is a default index of fidelity**

Adapted from Cordray et al.



Treatment Strength

Outcome

.45 .40 .35 .30 .25 .20 .15 .10 .05 .00

100 90 85 80 75 70 65 60 55 50

T^{Tx} → $\bar{Y}_T$

Infidelity

t^{tx} → $\bar{Y}_t$

Achieved Relative Strength =.15

t^{c} → $\bar{Y}_c$

"Infidelity"

T^{C} → $\bar{Y}_C$

(85)-(70) = 15

$$d = \frac{85-70}{30} = 0.50$$

$$d_{\text{with fidelity}} = \frac{\bar{Y}_T - \bar{Y}_C}{sd_{pooled}}$$

$$d_{\text{with fidelity}} = \frac{90-65}{30} = 0.83$$

$$d = \frac{\bar{Y}_t - \bar{Y}_c}{sd_{pooled}}$$

$$d = 0.50$$

Expected Relative Strength = (0.40-0.15) = 0.25

MAP ACADEMY

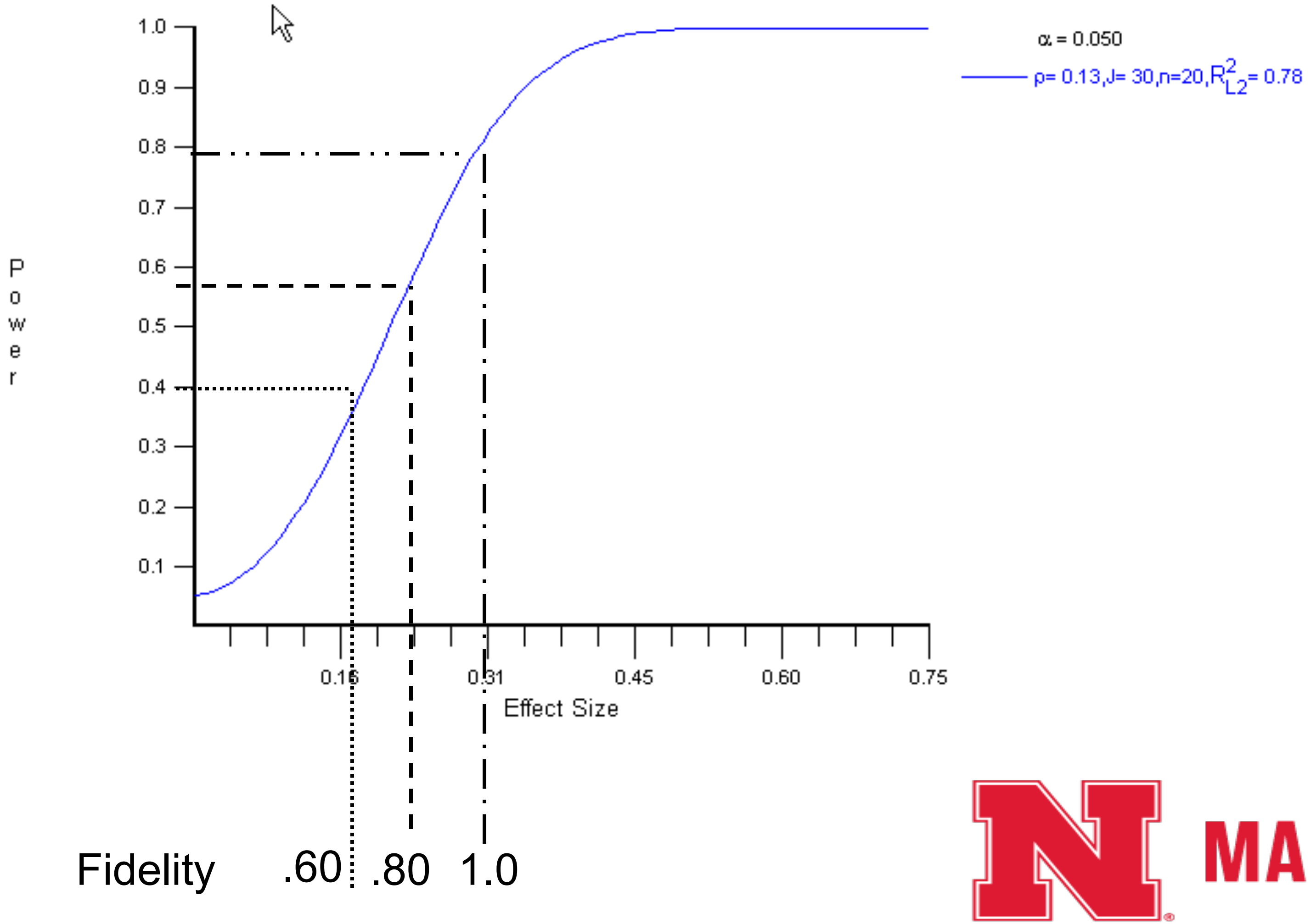
Why is this Important?

- **Statistical Conclusion validity**
  - **Unreliability of Treatment Implementation:** Variations across participants in the delivery receipt of the causal variable (e.g., treatment). Increases error and reduces the size of the effect; decreases chances of detecting covariation.
- Resulting in a reduction in statistical power or the need for a larger study….

Treatment Non-Compliance

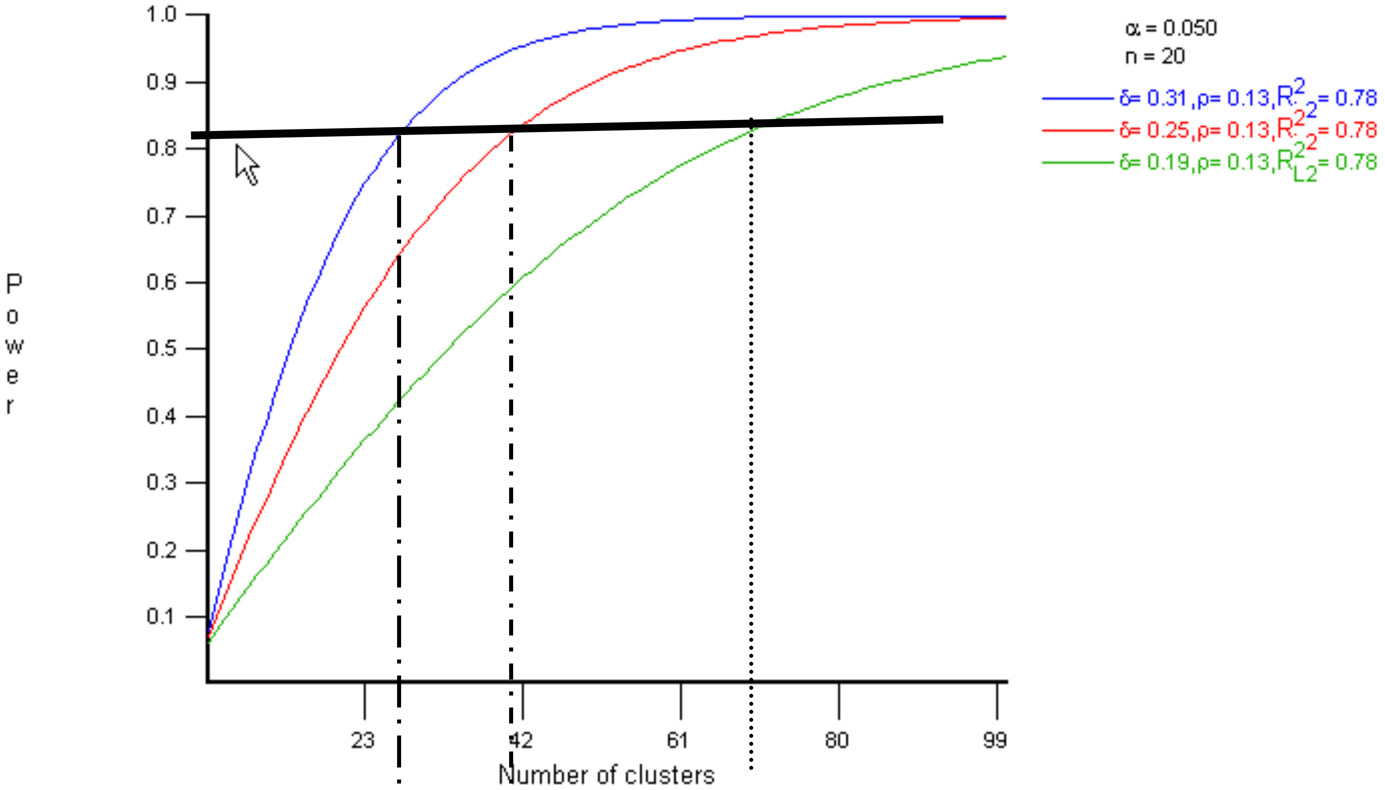
- It threatens the validity of an RCT
- It occurs
  1. when students assigned to treatment refuse to participate (drop-out)
  2. When those assigned to the control condition seek out and become members of the treatment group (drop-in)
- For Example:
  - 71% (17 of 24) of invited students in Year 1 and 59% (22 of 37) of students in Year 2 chose to attend
    - → the overall non-compliance rate: 36% (22 of 61)
  - Days of summer school attendance: range from 5 to 20(i.e. Perfect attendance) and overall mean of 17.21 (SD= 2.97).
    - The imperfect attendance rate → some summer school students did not receive a full dose of supplemental instruction

The Effects Structural Infidelity on Power



MAP ACADEMY

Influence of Infidelity on Study-size



Fidelity 1.0 .80 .60

MAP ACADEMY

If That Isn’t Enough….

- **Construct Validity:**
  - **Which is the cause? (T^{Tx} - T^{C}) or (t^{Tx} – t^{C})**
    - **Poor implementation:** essential elements of the treatment are incompletely implemented.
    - **Contamination:** The essential elements of the treatment group are found in the control condition (to varying degrees).
    - **Pre-existing similarities between T and C on intervention components.**
- **External validity – generalization is about (t^{Tx} - t^{C})**
  - **This difference needs to be known for proper generalization and future specification of the intervention components**

MAP ACADEMY

NEBRASKA ACADEMY FOR
METHODOLOGY, ANALYTICS & PSYCHOMETRICS

Measuring and Calculating Fidelity

University of Nebraska Lincoln

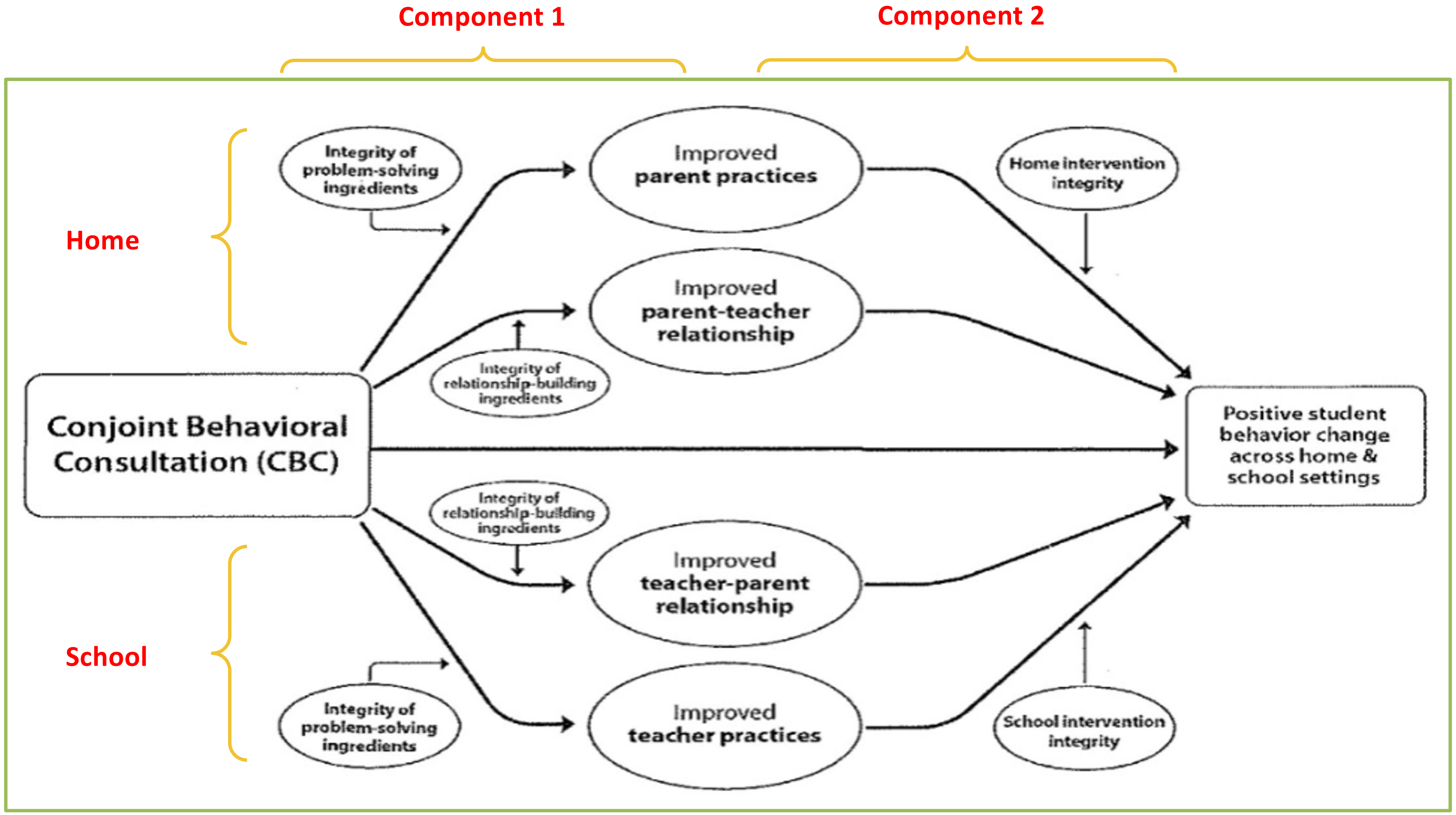
In Practice….

- Identify core components in the intervention group
  - e.g., via a Model of Change
- Establish benchmarks (if possible) for T^{TX} and T^{C}
- Measure core components to derive t^{Tx} and t^{C}
  - e.g., via a “Logic model” based on Model of Change
- Measurement (deriving indicators)
- Convert to Achieved Relative Strength and implementation fidelity scales
- Incorporate into the analysis of effects

Specifying Intervention Models

- Simple version of the question: ***What was intended?***
- Interventions are generally multi-component, sequences of actions
- Mature-enough interventions are specifiable as:
  - Conceptual model of change
  - Intervention-specific model
  - Context-specific model

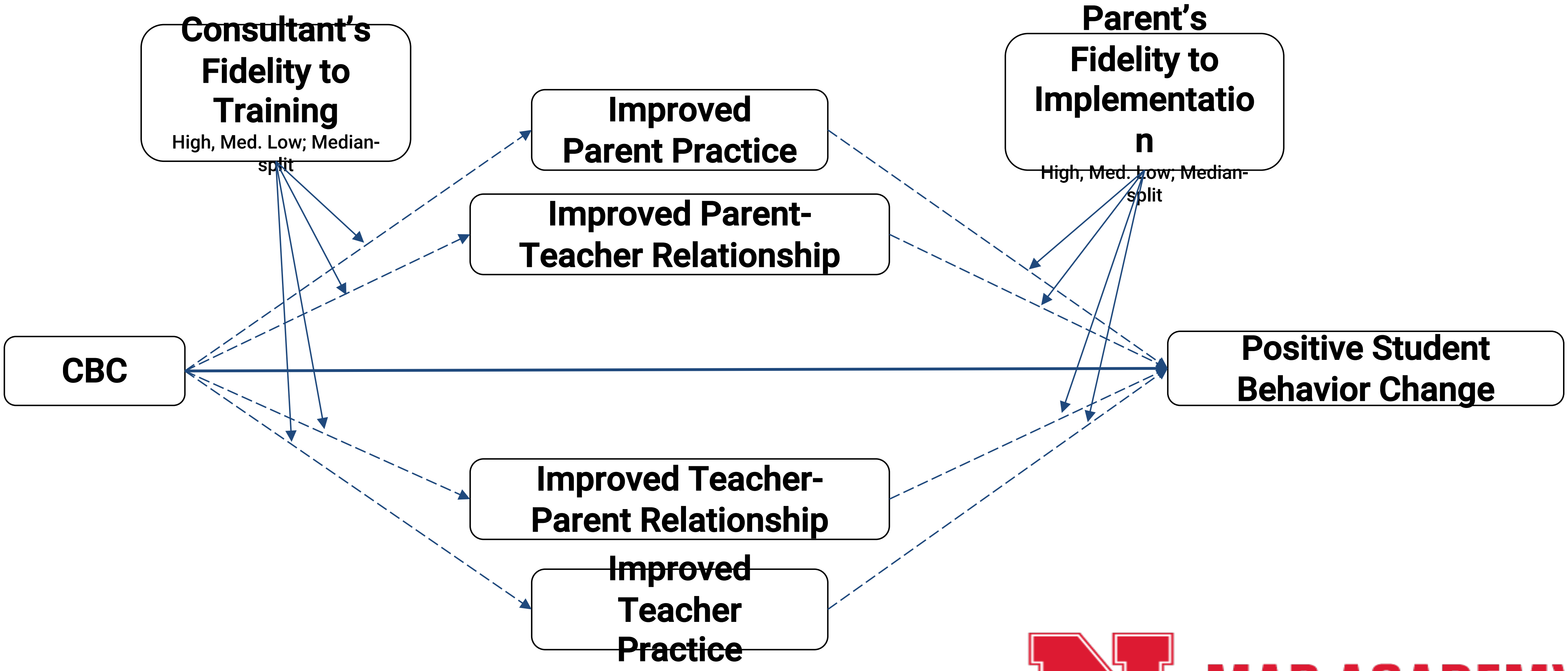
Define the Components of the Intervention



Component 1
Component 2
Home
School
Integrity of problem-solving ingredients
Improved parent practices
Home intervention integrity
Improved parent-teacher relationship
Integrity of relationship-building ingredients
Conjoint Behavioral Consultation (CBC)
Positive student behavior change across home & school settings
Integrity of relationship-building ingredients
Improved teacher-parent relationship
Integrity of problem-solving ingredients
Improved teacher practices
School intervention integrity

MAP ACADEMY

Determine Where & How Fidelity Plays a Role



Consultant's Fidelity to Training

High, Med. Low; Median-split

Parent's Fidelity to Implementation

High, Med. Low; Median-split

Improved Parent Practice

Improved Parent-Teacher Relationship

CBC

Positive Student Behavior Change

Improved Teacher-Parent Relationship

Improved Teacher Practice

MAP ACADEMY

What do we measure?

What are the options?

(1) Essential or ***core*** components (activities, processes);

(2) Necessary, but not unique, activities, processes and structures (supporting the essential components of T); and

(3) Ordinary features of the setting (shared with the control group)

- Focus on 1 and 2.

Quality Measures of Core Components

- Measures of resources, activities, outputs
- Range from simple counts to sophisticated scaling of constructs
- Generally, involves multiple methods
- Multiple indicators for each major component/activity
- Reliable scales (3-4 items per sub-scale)

Initial Development of an Intervention-Specific Fidelity Measure

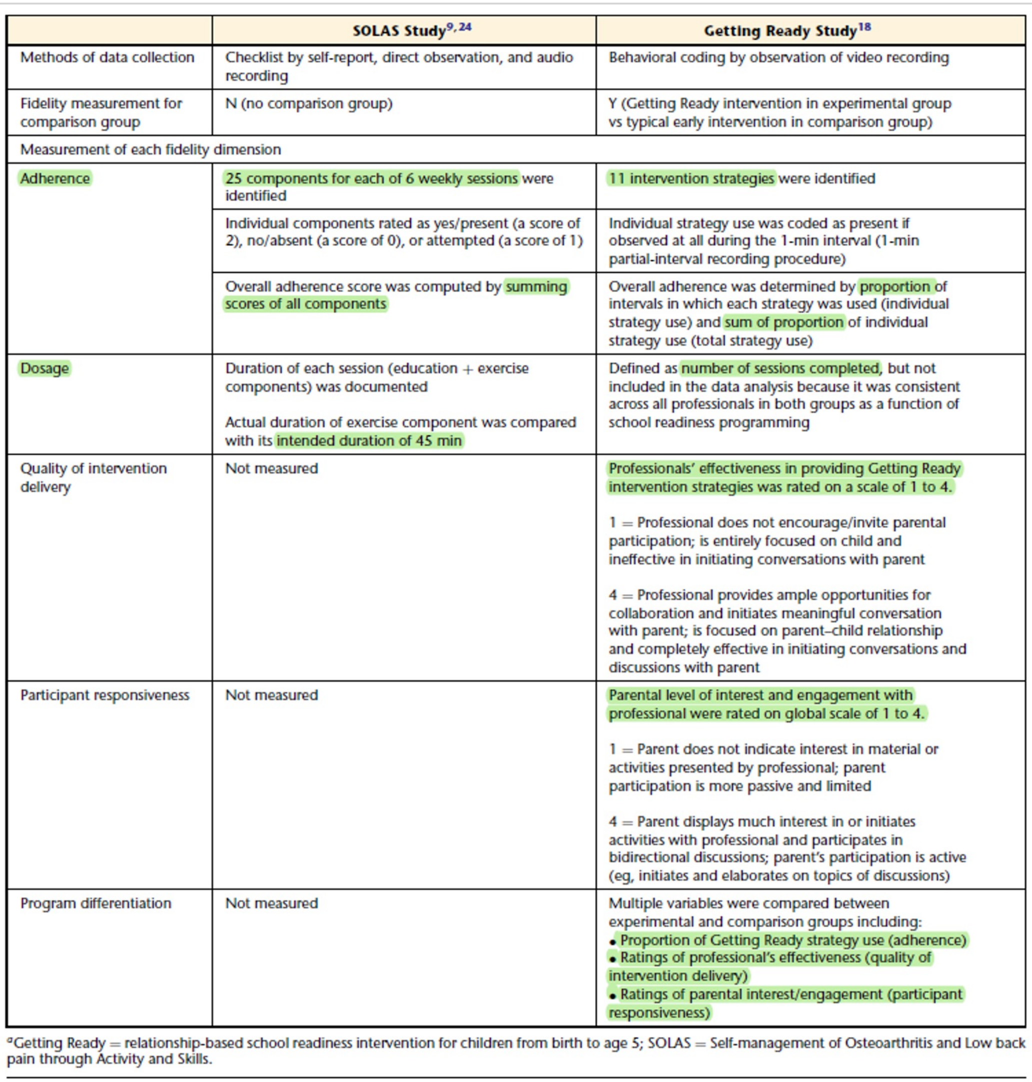
1. Identify possible indicators or key components of the approach by either expert consensus or previous research
2. Establish a measurement system, which involves decisions about how to measure the key components (eg, direct observation, coding video, or use of products such as written notes from the intervention) and how to determine if the intervention is implemented with acceptable fidelity.
3. Examine reliability and validity of data from the the fidelity measurement instrument.
   - percentage of agreement, coefficient kappa, intraclass correlation coefficient, or Pearson correlation coefficients.
   - Group method:examining differences in fidelity scores across interventions.
   - Convergent validity: examining the agreement b/t 2 different sources of information

Common Methods of Measuring Fidelity

- **Self-Report** - assesses adherence to the implementation of the intervention from the interventionist's perspective computed as a percentage of steps completed by the participants
  - + lessened reliance on extra human material resources
  - - Some suggest this results in the overestimation of implementation
- **Permanent Product** – assesses adherence to the implementation of the intervention through tangible evidence generated on intervention records/protocols (ie. charts, tokens, or home-school notes)
  - + offer relatively simple measurement procedures without additional work from the interventionist
  - - This doesn't work when trying to understand behavior compliance, quality of work, or appropriateness of social responses
- **Direct Observation** – assesses adherence to the implementation of the intervention involves a trained individual assessing objective implementation of treatment plan
  - + can be modified to address any intervention
  - - Very resource intensive

MAP ACADEMY

Develop a Concept Map of Fidelity



	SOLAS Study[9,24]	Getting Ready Study[18]
Methods of data collection	Checklist by self-report, direct observation, and audio recording	Behavioral coding by observation of video recording
Fidelity measurement for comparison group	N (no comparison group)	Y (Getting Ready intervention in experimental group vs typical early intervention in comparison group)
Measurement of each fidelity dimension		
Adherence	25 components for each of 6 weekly sessions were identified	11 intervention strategies were identified
	Individual components rated as yes/present (a score of 2), no/absent (a score of 0), or attempted (a score of 1)	Individual strategy use was coded as present if observed at all during the 1-min interval (1-min partial-interval recording procedure)
	Overall adherence score was computed by summing scores of all components	Overall adherence was determined by proportion of intervals in which each strategy was used (individual strategy use) and sum of proportion of individual strategy use (total strategy use)
Dosage	Duration of each session (education + exercise components) was documented Actual duration of exercise component was compared with its intended duration of 45 min	Defined as number of sessions completed, but not included in the data analysis because it was consistent across all professionals in both groups as a function of school readiness programming
Quality of intervention delivery	Not measured	Professionals' effectiveness in providing Getting Ready intervention strategies was rated on a scale of 1 to 4. 1 = Professional does not encourage/invite parental participation; is entirely focused on child and ineffective in initiating conversations with parent 4 = Professional provides ample opportunities for collaboration and initiates meaningful conversation with parent; is focused on parent–child relationship and completely effective in initiating conversations and discussions with parent
Participant responsiveness	Not measured	Parental level of interest and engagement with professional were rated on global scale of 1 to 4. 1 = Parent does not indicate interest in material or activities presented by professional; parent participation is more passive and limited 4 = Parent displays much interest in or initiates activities with professional and participates in bidirectional discussions; parent's participation is active (eg, initiates and elaborates on topics of discussions)
Program differentiation	Not measured	Multiple variables were compared between experimental and comparison groups including: • Proportion of Getting Ready strategy use (adherence) • Ratings of professional's effectiveness (quality of intervention delivery) • Ratings of parental interest/engagement (participant responsiveness)

[a]Getting Ready = relationship-based school readiness intervention for children from birth to age 5; SOLAS = Self-management of Osteoarthritis and Low back pain through Activity and Skills.

MAP ACADEMY

Develop a Concept Map of Fidelity

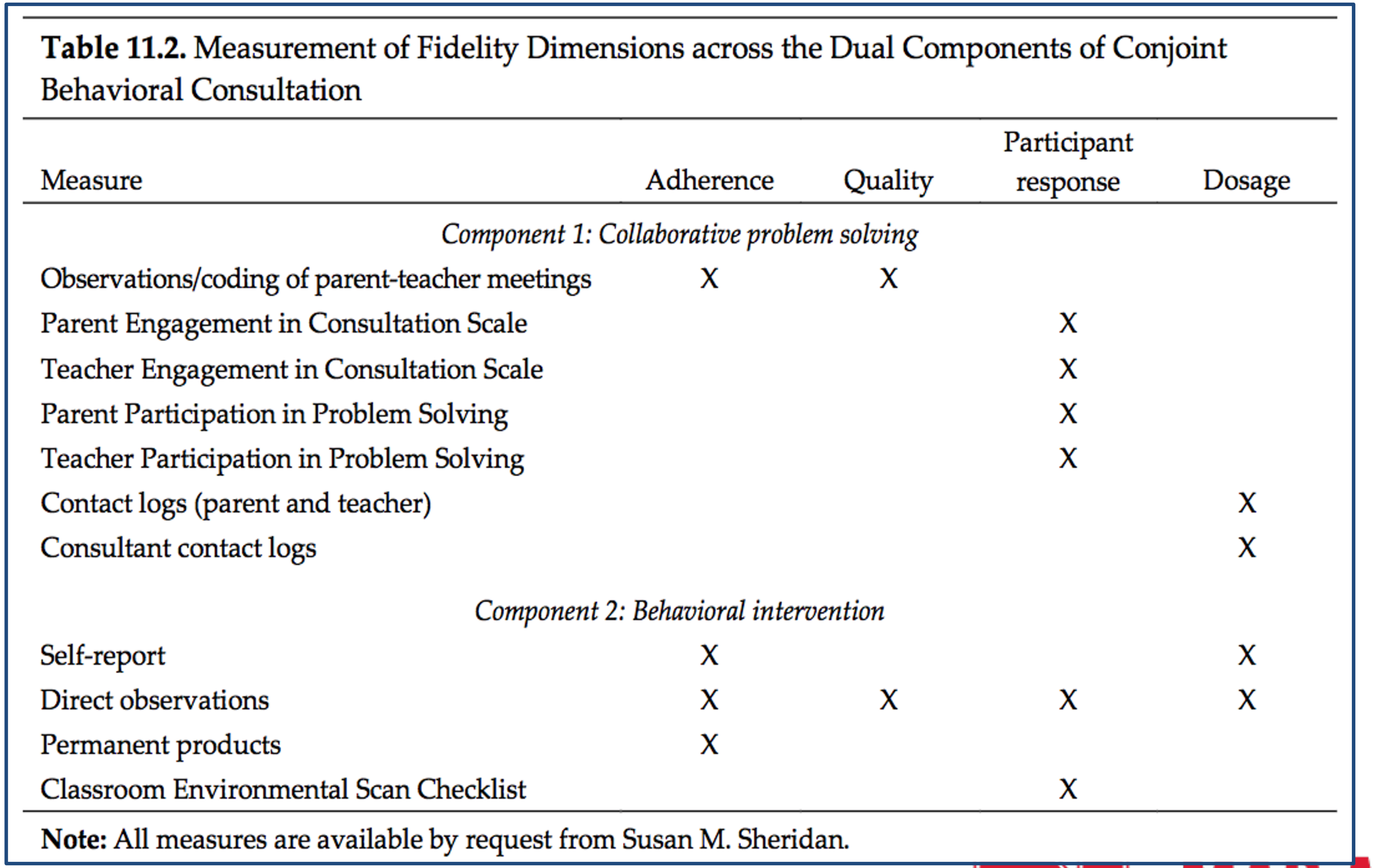


Table 11.2. Measurement of Fidelity Dimensions across the Dual Components of Conjoint Behavioral Consultation

Measure	Adherence	Quality	Participant response	Dosage
Component 1: Collaborative problem solving				
Observations/coding of parent-teacher meetings	X	X		
Parent Engagement in Consultation Scale			X	
Teacher Engagement in Consultation Scale			X	
Parent Participation in Problem Solving			X	
Teacher Participation in Problem Solving			X	
Contact logs (parent and teacher)				X
Consultant contact logs				X
Component 2: Behavioral intervention				
Self-report	X			X
Direct observations	X	X	X	X
Permanent products	X			
Classroom Environmental Scan Checklist			X	

Note: All measures are available by request from Susan M. Sheridan.

Reading First Implementation: Specifying Components and Operationalization

Components	Sub-components	Facets	Indicators (I/F)
Reading Instruction	Instructional Time	2	2 (1)
	Instructional Materials	4	12 (3)
	Instructional Activities /Strategies	8	28 (3.5)
Support for Struggling Readers (SR)	Intervention Services	3	12 (4)
	Supports for Struggling Readers	2	16 (8)
	Supports for ELL/SPED	2	5 (2.5)
Assessment	Selection/Interpretation	5	12 (2.4)
	Types of Assessment	3	9 (3)
	Use by Teachers	1	7 (7)
Professional development	Improved Reading Instruction	11	67 (6.1)
4	**10**	**41**	**170 (4)**

Adapted from Moss et al. 2008

Reading First Implementation: Some Results

Components	**Sub-components**	**Performance Levels**	
		RF	Non-RF
Reading Instruction	Instructional Time (minutes)	101	78
	Support	79%	58%
Struggling Readers	More Tx, Time, Supplemental Service	83%	74%
Professional Development	Hours of PD	41.5	17.6
	Five reading dimensions	86%	62%
Assessment	Grouping, progress, needs	84%	71%

Adapted from Moss et al. 2008

MAP ACADEMY

Analytic Procedures

ITT analysis:

- ITT: A policy-relevant causal estimate of the effect of treatment assignment
- It does not reflect the effect of treatment receipt without perfect compliance

AT model:

- An estimate of the treatment effect for the subgroup of summer school students who complied with their assignment.

ITT+weighting methods:

- To correct for potential bias in the estimation of treatment effects from non-compliance with the treatment offer
- A compliance-adjusted treatment receipt estimate: by weighting the ITT by the proportion of treatment students who actually attended summer school (M adjusted- M control/P compilers, P = the proportion of compilers)

Instrumental Variable (IV) analysis:

- An approach to account for treatment non-compliance
- ITT estimate/proportion of compilers or a two-stage least-square regression model.

Model-based Complier Average Causal Effect (CACE) methods:

- An estimate of the local effect of treatment recept
- Unbiased estimation of the effects of an intervention by modeling unknown compliance status as missing data

Distinguishing Implementation Assessment from the Assessment of Implementation Fidelity

- Two ends on a continuum of intervention implementation/fidelity:
- A ***purely* descriptive** model:
  - Answering the question “What transpired as the intervention was put in place (implemented).
- Based on ***a priori* intervention model**, with explicit expectations about implementation of program components:
  - Fidelity is the extent to which the realized intervention ($\mathbf{t}^{Tx}$) is faithful to the ***pre-stated*** intervention model ($\mathbf{T}^{Tx}$ **)**
  - Infidelity = $\mathbf{T}^{Tx} - \mathbf{t}^{Tx}$
- Most implementation fidelity assessments involve descriptive and model-based approaches.

MAP ACADEMY

Incorporate into Statistical Analysis

$$Y_{ij} = \beta_{0j} + \beta_{1j} \text{ (Assignment Status) } + r_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{ (Mean Instructional Hours) } + u_{0j} \quad (2a)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \text{ (Treatment Contrast) } + u_{1j} \quad (2b)$$

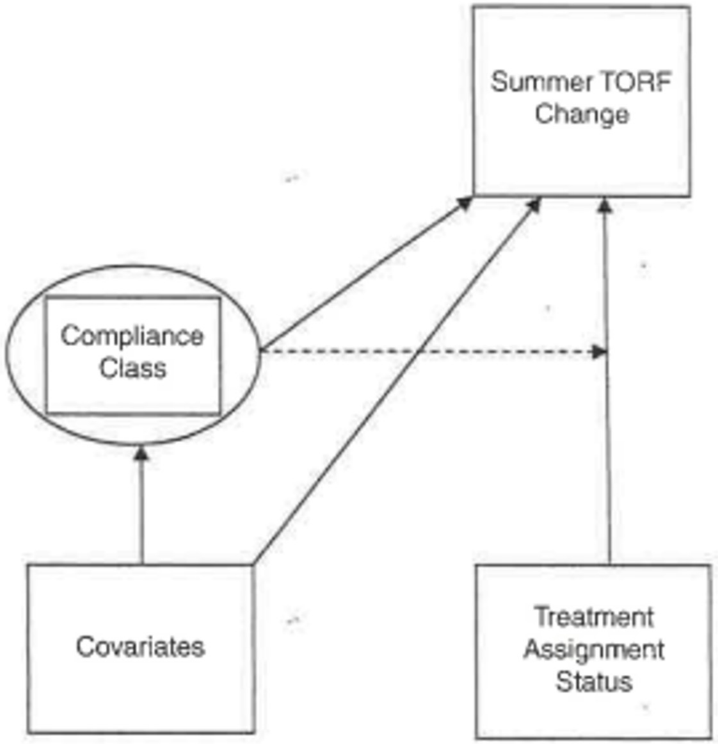


FIGURE 5.1 Summer School CACE Model.

Thank you!

jbovaird2@unl.edu
http://mapacademy.unl.edu
https://cehs.unl.edu/edpsych/faculty/james-bovaird

Nebraska
UNIVERSITY OF
Lincoln