How Can We Determine Whether We Have a Good Measure of Trauma?

Terry L Schell
RAND, Santa Monica

(The views expressed in this presentation do not reflect the views of RAND)

Overview

- Brief review of standard psychometric theory
- Why the assumptions of these theories, and the related techniques, should not be applied to measures of trauma
- Implications for
  - How to score trauma scales
  - How to develop trauma scales
  - How to evaluate the performance of trauma scales

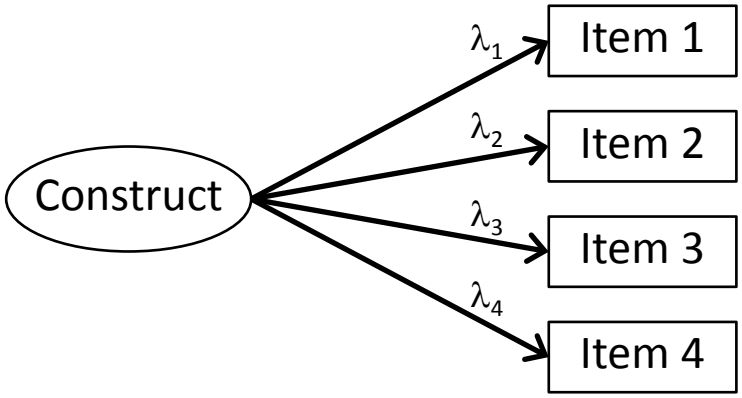
Psychometric theory of reflexive measures

- *Reflexive*, or *effect-indicated*, measures have items that are theorized to share a common cause
- This common cause is the construct to be measured
  - Items "reflect" the influence of the construct
  - Observed indicators are the "effect" of the construct
- This causal assumption is the basis of almost all psychometric analyses, including:
  - Classical test theory
  - Factor analysis
  - Item response theory
- Most standard psychological measures are reflexive

Reflexive measures have correlated items

Items in a reflexive measure are correlated due to their shared cause.

- Items are best when they are strongly caused by the construct
- Correlation between items 1 and 2 is proportional to $\lambda_1\lambda_2$
- The quality of the measurement can be inferred from the correlation between items



Summing correlated items converges on an error-free measure of the common cause

- Variance of the sum of two random variables:

$$var_{sum} = var_1 + var_2 + 2(cov_{12})$$

- Because of the covariance term, the sum of the two items is more associated with the common cause than either item.
- With more items the covariance terms dominate:

$$var_{sum} = var_1 + var_2 + var_3 + var_4 + 2(cov_{12} + cov_{13} + cov_{14} + cov_{23} + cov_{24} + cov_{34})$$

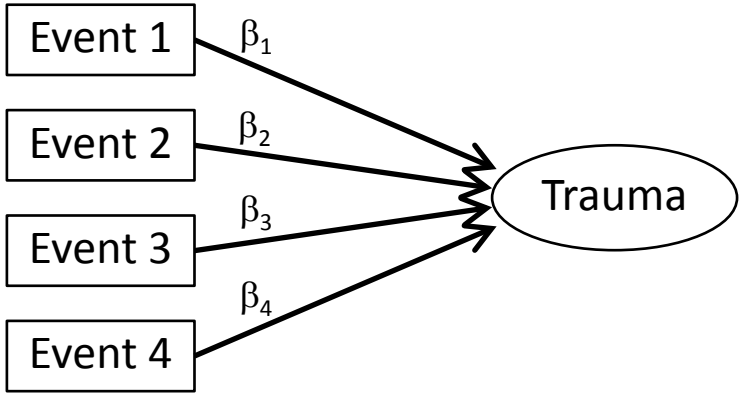
- The sum of items converges on an error-free measure of the common cause as
  - the number of items goes to infinity, and
  - The correlation between items goes toward 1
- Cronbach's α is a measure of the extent to which the covariance terms dominate in the variance of the sum, and is a simple function of the number of items and the average correlation between them.

Psychometric applications for reflexive scales

- Advanced psychometric methods will allow a the measure to converge to error-free more quickly than a simple sum of items:
  - Items that are more correlated with the other items can be given more weight in the sum (e.g., PCA, factor analysis, IRT)
  - The portion of variance that appears to be unique to an item -- not caused by the common cause -- can be subtracted out of the scale (e.g., factor analysis, IRT)
- An error-free measure of the hypothetical common cause of the items is not always an error-free measure of the intended construct
  - The causal model could be wrong
  - There may be multiple shared causes, some of which you may not intend to measure, e.g., response biases, DIF
  - A reliable measure is not always valid

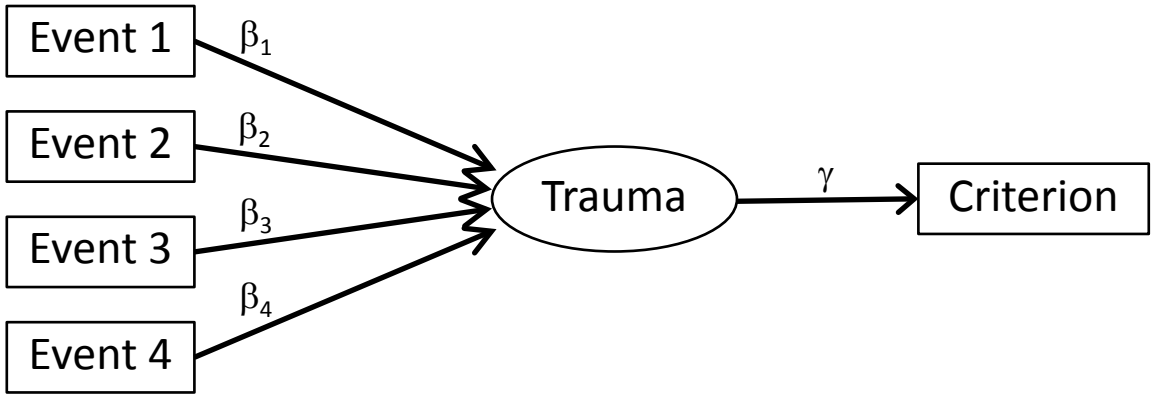
Trauma is not a reflexive construct

- We are not trying to measure the common cause of a set of life events
- Events are defined as traumatic because of their similar effects, not because they all share a single cause
- Trauma is a type of construct referred to as a:
  - *Formative* construct
  - *Cause-indicated* construct
  - *Composite* construct
- Traumatic events may or may not have a shared etiology, may or may not be correlated



How do you score a formative scale?

- Summing items creates a measure of the common cause, and should not be used unless the events are uncorrelated and equally predictive of the defined outcome
- The measure cannot be computed without some way to estimate the β's; unlike the λ's they cannot be inferred from correlations among the items
- A measurement criterion is necessary
  - Ideally, the study has included a measure of the effect(s) of trauma as specified by the theory
  - In such cases, the β weights can be estimated via regression of the criterion on the items



Features of using formative scales

- Using different criterion can result in different measures of "trauma," even for the same items.
  - The scoring of "trauma" for predicting PTSD symptoms may be different than when predicting drug use, domestic violence or other outcomes.
  - The concept of unidimensionality is undefined
  - Each item is intended to have "unique" variance that is not error
- Because formative scales are defined by theoretically defined predictive criteria, focus is generally on validity not reliability.
  - Internal consistency reliability is not defined
  - Test-retest and other types of reliability are usually not assessed

Comparing reflexive and formative scales

Unlike reflexive scales, formative scales are…

- optimally efficient when items/events are uncorrelated
- less valid with higher values of Cronbach's α when they have been scored as summed scales
- less influenced by items that are highly correlated with other items, rather than more influenced
- generally problematic outcomes.
  - The scale may not be well predicted by other variables, even when the individual items/events are well-predicted
  - Usually need to be broken apart when treated as outcomes

Alternative scoring for formative scales?

- Score according to strict construct definitions when available
  - Legal definition of sexual harassment
  - Prior research may exist that estimates or approximates the β's
- Combine events without summing
  - Did any of the specified events occur?
- Minimize covariance before summing
  - Drop/combine redundant items
  - Give redundant items less weight
- Even when you have measured the theorized criterion, it may be desirable to use a unit-weighted summed scale
  - This can be done whenever the set of β's can be set to equal without significant loss of predictive power
  - Large trauma scales can often be replaced with shorter subsets

Measurement equivalence examples

- Do the β values vary across gender, racial, or geographic groups?
- Does the γ value vary across gender, racial, or geographic groups?
- Does the measure attenuate the zero order relationship between PTSD symptoms and those demographic variables not theorized to be causally related to PTSD?
- Unlike with reflexive scales, there is no definition of item bias. The items are fully exogenous and could have any relationship with the demographic characteristics