MPLUS AND THE APPROPRIATE ANALYSIS OF ORDINAL SCALES AND HIERARCHICAL DATA IN SOCIAL WORK RESEARCH

Natasha K. Bowen, UNC-CH
Kate M. Wegmann, UNC-CH
Main Points

1. It is best to analyze data collected with multiple scale items in the structural equation modeling framework
Main Points

2. Special procedures are needed to analyze ORDINAL scale data appropriately
3. The hierarchical, or clustered, nature of data should be addressed in analysis procedures
4. Mplus software easily and appropriately handles clustered, ordinal–level scale data (as well as missing values and weights) in confirmatory factor analyses and general SEM models.
Resources

- References, this presentation, and a detailed document with instructions on using Mplus for CFA and general SEM will be posted online at the following website for the SSWR school research special interest group:

- [http://www.luc.edu/sswsig/](http://www.luc.edu/sswsig/)

“Resources for Researchers” link
Disclosure

- No financial interest in Mplus
- Sharing experiences of self and students with the program
- Backed by literature
- Mplus does a lot more than discussed here
- Developers may make other recommendations
Plan

- Nature of social work data
- Analysis issues
- SEM with Mplus
  - Solutions
  - How to
- Your questions, experiences, expertise at any time
Starting Points

Regression

EFA

Programs used

Options used

CFA

Programs used
Common Sources of Social Work Data

- Administrative datasets
- National and other large-scale survey datasets
- Practice assessments
- Instruments developed by researchers
Measurement Level of Social Work Data

- Nominal
- Ordinal
- Interval-level or Continuous

Sets of variables to measure constructs

Why so many ordinal measures?
Other characteristics of SW data

- Clustered
- Weighted
- Missing
Main Point 1

Analyzing sets of variables that measure the same construct (scales)
Analyzing scales

- Set of Likert scaled items assessing symptoms of depression
- Several items gleaned from national data set that seem related to a construct of interest to the researcher
- Quality of life measure a researcher developed for residents of assisted living center
Common approaches to item sets

- Put individual items into regression models

Difficulty modeling multiple highly correlated predictors
  - multicollinearity
  - difficulty interpreting results
Common approaches to item sets

**Composites**
sum or average items’ values into one score

**Exploratory Factor Analysis**
identify which items “hang” together, how many factors, nature of factors, weights of different items
Limitations of common approaches

**Composites**
- Also sum or average measurement error in raw scores
- Usually equal weights assumed
- Reliability issues

**Exploratory Factor Analysis**
- Principal components
- Principal axis
- Maximum likelihood
- Limited modeling flexibility
- Limited model testing and comparing
Advantages of CFA

Flexible modeling

- Some factors can be correlated, others not
- Items can load on any number of factors (usually one or two)
- Second order models possible
Advantages of CFA

• Examines only construct-related variance of indicators
• More model tests and comparisons
• More rigorous test of scales
• Can conduct multiple group tests
• Basis for latent variable modeling in structural models
Main Point 1—Case Closed?

1. It is best to analyze data collected with multiple scale items in the structural equation modeling framework.
Special procedures are needed to analyze ORDINAL scale data appropriately.
Core issue with ordinal data:

Values assigned to response options are arbitrary and lack quantitative meaning.
<table>
<thead>
<tr>
<th>Never</th>
<th>Sometimes</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Lack Quantitative Meaning

Never  rarely  Sometimes  often  Always

1  2  3  4  5

Clear from the arbitrary trait, but also:
- no quantitative units associated with anchors
- can’t assign numbers to spacing between units
- even a simple mean is meaningless numerically
Common approach with Ordinal Data

- Analyze them as if they were interval level or higher
  - Means tests
  - EFA’s
  - Regressions
“Ordinal variables are not continuous variables and should not be treated as if they are. It’s common practice to treat scores 1, 2, 3 . . . . assigned to categories as if they have metric properties, but this is WRONG. Ordinal variables do not have origins or units of measurement. Means, variances, and covariances of ordinal variables have NO MEANING.”

Ordinal Data

Associated issues:

1. Non-normality of variable distributions
Normality Unlikely with Ordinality

Yes No
Dissatisfied Neutral Satisfied

1 5 10 15 20 25 30 35 40 45 50 55 60 65 70
Non-Normality

Violation of assumptions of analysis procedures that can affect estimates and significance tests
Common approaches

Variable Transformations
- Square, invert, take square root or log . . .

Use Robust Estimator
- General Least Squares or Principal Axis in EFA
- MLR, WLS in CFA
Limitations of common approaches

Variable Transformations

- Guidelines about when to use not clear
- Each variable in a question set may require a different transformation
- Results may be difficult to interpret

Use Robust Estimator

- Guidelines about when to use not clear
  - how much non-normality?
Limitations of common approaches when variables are ordinal
The hierarchical, or clustered, nature of data should be addressed in analysis procedures.
Observations are independent (error terms of individual respondents are not correlated). This assumption is violated when:

- data come from students in the same classroom, patients from the same clinic, families in the same neighborhood
- staged sampling strategy started with higher level units (e.g., states) and continued to lower level units (e.g., cities in states)
- data come from same subjects over time
Consequence of Violation

- Underestimated standard errors, which may lead to Type I errors — finding a significant effect when there is no true effect (false positive)
Example

Taking account of clustering

- Estimate: 85.6328
  - SE: 0.7148
  - DF: 23
  - t value: 119.80

Not taking account of clustering

- Estimate: 85.5901
  - SE: 0.3194
  - DF: 520
  - t value: 267.99
Common approaches

Ignore clustering

Not good

Report a low inter-class correlation and proceed

How low is low enough?

Use Hierarchical Linear Modeling

Commendable but how to do this when variables are scales and/or ordinal?
4. Mplus software easily and appropriately handles clustered, ordinal–level scale data (as well as missing values and weights) in confirmatory factor analyses and general SEM models.
Mplus solutions

Structural equation modeling program

- Can analyze data of all measurement levels and distributions
- Can include multiple cluster variables
- Full information maximum likelihood (FIML) approach to missing values on dependent variables (such as, indicators in CFA and latent variable structural models)
CFA modeling of scales
- Unobserved latent phenomena drive scores on hypothesized indicators of constructs.
- Shared variance among hypothesized indicators represents the construct of interest.
Mplus and Main Point 1

- Unshared variance of indicators, including systematic and random measurement error is separated from latent variables

We want to use this approach to the analysis of data from scales
Mplus and Main Point 2

Theoretical approach to Categorical and Ordinal Data

![Histogram with frequency counts and statistical measures: Mean = 3.4, Std. Dev. = 0.933, N = 424.](image)
Special correlation matrices

Takes into account core issue with ordinal data:

Values assigned to ranked categories have no true quantitative meaning
Special correlation matrices

Assumes underlying normal distribution behind the cut points on an ordinal item.
Special correlation matrices

Assumes underlying normal distribution behind the cut points on an ordinal item

Dissatisfied 1 | Neutral 2 | Satisfied 3

2 thresholds
Special correlation matrices

All 2-way contingency tables used to create correlation matrix.

Polyserial
Polychoric
Tetrachoric

Correlation matrix analyzed with Weighted Least Squares Estimator (WLS).
Mplus and Main Point 2

All of the above happens when variables are designated as categorical (nominal)

WLSMV is recommended by program developers and in the literature

We want to use this approach to the analysis of data from ordinal scales
We can conduct our CFA or SEM analysis with properly modeled scale data in hierarchical models with two or three levels.

(LONGITUDINAL MODELS WOULD BE HANDLED WITH GROWTH CURVES)

For example:

Patients in hospitals at Level 1
Hospitals at Level 2

Patients at Level 1
Hospital units at Level 2
Hospitals at Level 3
One higher order grouping variable can be modeled simply to correct for underestimation of standard errors (COMPLEX), or variables on two higher order units can be included as predictors (TWOLEVEL).

For example:

Patients—Level 1
Hospital—Level 2

Hospital level predictors—staff training, staff satisfaction, etc.
We want to use this approach to the analysis of data from scales when subjects are clustered in higher level units.
Using Mplus

Raw data file preparation

- All data cleaning needs to happen in your general statistics or spreadsheet file before analysis in Mplus
- Create any new or recoded variables
- Designate missing values
  - non-numeric flags; only one per dataset
  - numeric flags; you can have different ones for every variable, one for the whole dataset; be sure flags don’t overlap with valid response options
Using Mplus

Raw data file preparation

- Before moving to Mplus, be sure you know the names of your variables and order in which they appear in the dataset.

- Save saved as free or fixed ASCII files (tab, comma, or space delimited) (free format easiest).

- Mplus can read tab delimited .dta files saved in SPSS as well as text MS-DOS files (.txt) or "formatted text, space delimited" files (.prn) saved in Microsoft Excel.

- Do not save variable names!
Using Mplus

Matrix file preparation

- lower triangle or full covariance matrix: free file format, matrix begins on first line of the file
- correlation matrix: also free file format, but first line is the means of variables, standard deviations on the second line, and first line of the matrix (either full or lower triangle) on the third line
Using Mplus

Text based
10 basic commands
We’ll use:

TITLE
DATA
VARIABLE
ANALYSIS
MODEL
OUTPUT
SAVEDATA

Command statements end in semi-colon
Using Mplus

**TITLE**

Optional

Will appear in output

Need to update after each model change
Using Mplus

DATA

Tell Mplus where the data file is and what kind of data file it is (TYPE IS)

Give full directory and file name with extension

Default is raw data file; no need to specify TYPE
Using Mplus

AGGRAVATION SAVERS:
If there are blanks in the directory label, put entire directory inside quotation marks
Must have extension on file name
VARIABLE names should have 8 or fewer characters
Input lines cannot exceed 80 characters
(hit return)
ERRORS aren’t necessarily fatal
Using Mplus

VARIABLE

Important command for us, because our ordinal scale data are different from the default

Need to

- list names of all variables in dataset (8chars)
- indicate which variables will be used in current analysis
- indicate how missing values are designated
- indicated which variables are cluster variables
- indicate which variables are nominal and categorical
Using Mplus

ANALYSIS COMMAND

Details on the analysis:

- estimator
- how to handle missing values
- clustering
- iteration criteria
- convergence criteria
Using Mplus

**MODEL**

In a CFA, this is where you specify names of latent variables and which items load on which factor.

Very simple language, especially if you work with the defaults (first item listed for a factor has its loading fixed at 1; all factors are assumed to be correlated; no correlated errors).

In structural model, this is where you specify relationships among factors.
Using Mplus

OUTPUT

At a minimum ask for standardized estimates
Also: sample statistics, modification indices
SAVEDATA

You can save factor scores, matrices generated from the analysis

Information for use in chi square difference test
Scale data best analyzed in SEM framework
Ordinal data most appropriately analyzed with polychoric correlation matrix and WLS(MV)
Clustering of data should not be ignored
Mplus does all this (plus FIML)

Social Work research needs to use these sophisticated and appropriate methods to be credible and marketable and to get most accurate results for translation into practice