Factor Analysis

MERMAID Series 12/11

Galen E. Switzer, PhD
Rachel Hess, MD, MS
Ways to Examine Groups of Things

- **Groups of People**
  - Cluster Analysis
  - Latent Class Analysis

- **Groups of Indicators**
  - Exploratory Factor Analysis
  - Confirmatory Factor Analysis (Structural Equation Modeling)
Groups of People vs Groups of Indicators/Items

**Groups of People**
- Group Learners
- Individual Learners
- Mixed Learners

**Groups of Indicators**
- Learning Style
- Mentoring Quality
- Life Stressors

Which set of learners is most successful in a traditional medical training environment?
Which is the strongest predictor of medical school success?
Latent Structure and Factor Analysis

- Expertise is an unobserved, latent construct/concept
- Expertise is measured with multiple indicators
- Expertise could be a single construct or a construct with subdomains
- These subdomains are also referred to as “factors”
SF-12 Example

Physical Dimension
- General health
- Physical limits on daily activities
- Physical limits on strenuous activities
- Physical limits on ability to accomplish things
- Physical limits on work activities
- Pain

Emotional Health Dimension
- Emotional limits on ability to accomplish things
- Emotional limits on work and other activities
- Calm and peaceful
- Energy
- Feeling down
- Interference with social activities
Systematic reviews indicate a lack of reporting of reliability and validity evidence in subsets of the medical education literature. A comprehensive review of exploratory factor analysis in instrument development across the continuum of medical education had not been previously identified. This study is a critical review of instrument development articles employing exploratory factor or principal component analysis published in medical education. Data extraction of 64 articles published in the peer-reviewed medical education literature indicates significant errors in the translation of exploratory factor analysis best practices to current practice. Instruments reviewed for this study lacked supporting evidence based on relationships with other variables and response process, and evidence based on consequences of testing was not evident. Findings suggest a need for further professional development within the medical education researcher community related to 1) appropriate factor analysis methodology and reporting and 2) the importance of pursuing multiple sources of reliability and validity evidence to construct a well-supported argument for the inferences made from the instrument. Medical education researchers and educators should be cautious in adopting instruments from the literature and carefully review available evidence. Finally, editors and reviewers are encouraged to recognize this gap in best practices and subsequently to promote instrument development research that is more consistent through the peer-review process.
Two Types of Factor Analysis

**Exploratory (EFA)**
- link between observed & latent variables unknown
  - Principal Components
  - Factor Analysis
    - max. likelihood factoring
    - alpha factoring
    - unweighted least squares
    - generalized least squares
  - Hybrids of FA/PCA
    - principal factors
    - image factoring

**Confirmatory (CFA)**
- link between observed & latent variables known/hypothesized
  - Factor Analysis Models
    - structural equation methods
Exploratory Factor Analysis

Specific Goals

• reduce the number of indicators/items required to describe a construct

• assess and improve a measure’s properties

• provide an operational definition for an underlying construct

• generate hypotheses
Exploratory Factor Analysis

Published Examples

• Differential behavioral patterns related to alcohol use in rodents: A factor analysis. (Salimov, 1999)

• Factor analysis of laboratory and clinical measurements of dyspnea in patients with chronic obstructive pulmonary disease. (Nguyen et al., 2003)

• Morphological abstraction of thyroid tumor cell nuclei using morphometry with factor analysis. (Murata et al., 2003)

• Factor analysis of the interrelationships between clinical variables in horses with colic. (Thoefner, et al., 2001)

• Factor analysis shows that female rat behavior is characterized primarily by activity, male rats are driven by sex and anxiety. (Fernandes et al., 1999)
EFA Terminology

- **Observed variable** – a variable measured directly
- **Latent variable** – an unobserved variable that “causes” observed variables
- **Correlation matrix** – a table showing all pairs of correlations between observed variables/indicators/items
- **Factor analysis** – a set of statistical procedures and judgments used to uncover and examine latent variables
- **Factor** – measurement term for a latent variable
- **Factor loading** – correlation between an indicator and an underlying factor – higher loading means the indicator better represents the factor
- **Eigenvalue** – a numeric value statistically assigned to each factor that corresponds to how much explanatory power that factor has
- **Scree plot** – a visual depiction of eigenvalues plotted by factor
- **Rotation** – a statistical procedure designed to find the “cleanest” possible factor solution – orthogonal or oblique
Observed vs Latent Variables

**Observed Variables**
- measured directly
- almost always contain error
- examples: body mass, lung volume, blood pressure, cell morphology

**Latent Variables**
- cannot be measured directly (unobserved)
- cause or produce observed variables
- examples: student satisfaction, academic success, family stress, burnout

![Diagram](attachment:image.png)
1. **Examine descriptive statistics**
   -- indicators on 1-5 scale

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Analysis N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind 1</td>
<td>2.5</td>
<td>0.40</td>
<td>309</td>
</tr>
<tr>
<td>Ind 2</td>
<td>5.1</td>
<td>0.38</td>
<td>318</td>
</tr>
<tr>
<td>Ind 3</td>
<td>4.2</td>
<td>3.90</td>
<td>317</td>
</tr>
</tbody>
</table>

- Data accurate?
- Indicators normally distributed?
- Missing data?
- Adequate sample size?  (good=300, very good=500, excellent=1,000)
2. Examine correlation matrix of variables

<table>
<thead>
<tr>
<th></th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
<th>Var 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind 1</td>
<td>1.00</td>
<td>0.40</td>
<td>0.15</td>
<td>1.00</td>
</tr>
<tr>
<td>Ind 2</td>
<td>0.27</td>
<td>0.30</td>
<td>0.68</td>
<td>...</td>
</tr>
</tbody>
</table>

- Variability in sizes of correlations?
- Matrix factorable? (several > .30)
- Outliers? (indicators that don’t correlate with any others)
EFA: Steps & Issues

3. Examine eigenvalues to decide how many factors to include

<table>
<thead>
<tr>
<th>Factor</th>
<th>Initial Eigenvalues</th>
<th>Total</th>
<th>% of Variance</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.211</td>
<td>43.48</td>
<td></td>
<td>43.48</td>
</tr>
<tr>
<td>2</td>
<td>1.784</td>
<td>12.49</td>
<td></td>
<td>55.97</td>
</tr>
<tr>
<td>3</td>
<td>0.856</td>
<td>5.99</td>
<td></td>
<td>61.97</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>N = 7 items</td>
<td></td>
<td>.</td>
<td></td>
<td>100.00</td>
</tr>
</tbody>
</table>

- Eigenvalue indicates the % of variance accounted for by the factor
- Provides first evidence of whether Factor Analysis is useful
- Generally identify factors with eigenvalues ≥ 1.0 as important
4. Examine scree plot to decide how many factors to extract

- factors are plotted by eigenvalue
- important factors are to the left of the scree
- tendency to over-extract using eigenvalues and under-extract using scree plot
EFA: Steps & Issues

5. Rotate the factor solution

Goal: clarify interpretation by maximizing high and minimizing low factor loadings

<table>
<thead>
<tr>
<th>Factor/Component</th>
<th>1</th>
<th>2</th>
<th>3 . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind 1</td>
<td>.45</td>
<td>.30</td>
<td>.12</td>
</tr>
<tr>
<td>Ind 2</td>
<td>.60</td>
<td>.14</td>
<td>.02</td>
</tr>
<tr>
<td>Ind 3</td>
<td>.01</td>
<td>.20</td>
<td>.20</td>
</tr>
<tr>
<td>Ind 4</td>
<td>.17</td>
<td>.20</td>
<td>.30</td>
</tr>
</tbody>
</table>

- loading = correlation between indicator and a given factor
- the matrix above is difficult to interpret
### Pain Example

<table>
<thead>
<tr>
<th>Pain Quality</th>
<th>Item</th>
<th>Promax Rotated Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Paroxysmal</td>
</tr>
<tr>
<td>Paroxysmal</td>
<td>Shooting</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td>Sharp</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>Electric</td>
<td>0.585</td>
</tr>
<tr>
<td></td>
<td>Hot</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>Radiating</td>
<td>0.363</td>
</tr>
<tr>
<td>Surface</td>
<td>Itchy</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>Cold</td>
<td>-0.104</td>
</tr>
<tr>
<td></td>
<td>Numb</td>
<td>-0.074</td>
</tr>
<tr>
<td></td>
<td>Sensitive</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>Tingling</td>
<td>0.227</td>
</tr>
<tr>
<td>Deep</td>
<td>Aching</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>Heavy</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>Dull</td>
<td>-0.178</td>
</tr>
<tr>
<td></td>
<td>Cramping</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Throbbing</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td>Tender</td>
<td>0.254</td>
</tr>
</tbody>
</table>

LBP indicates low back pain; OA, osteoarthritis; PQAS, Pain Quality Assessment Scale.

7. Determine if the factor solution is adequate
   • patterns should be evident in correlation matrix
   • two or more indicators per factor
   • indicator’s loading on its factor should exceed 0.30
   • “simple structure” has been achieved (i.e., fewer factors than indicators)
   • examine factors’ internal consistency reliability in separate alpha analysis
Summary: EFA Steps and Issues

**Exploratory Factor Analysis**

- examine descriptive statistics
- examine correlation matrix
- conduct EFA and examine eigenvalues
- examine scree plot
- select rotation type and rotate solution
- evaluate the adequacy of the final solution
Summary: Exploratory Factor Analysis

**Exploratory Factor Analysis**

- can be useful for empirically summarizing & reducing data
- does not involve statistical significance testing
- relies on guidelines rather than rigid rules
Confirmatory Factor Analysis

General Goals

• to test a hypothesized set of relationships among observed and unobserved variables

• to compare factor structure across samples

• to test construct/factor validity
CFA: Steps & Issues

1. Generate a visual model
   • depict relationships among observed and unobserved variables
   • depict relationships among unobserved variables
2. Evaluate Goodness of Fit Indices

- select appropriate indices
- GFI, AGFI, CFI etc. ≥ 0.90
- RMSEA ≤ 0.08
3. Evaluate alternate models
   - alternate models from EFA
   - alternate models suggested by CFA
   - compare goodness of fit indices
## CFA Fit Indices for Five Models

<table>
<thead>
<tr>
<th>Model</th>
<th>GFI</th>
<th>AGFI</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Factor</td>
<td>.937</td>
<td>.901</td>
<td>.859</td>
<td>.076</td>
</tr>
<tr>
<td>2 Factor Uncorrelated</td>
<td>.917</td>
<td>.870</td>
<td>.753</td>
<td>.101</td>
</tr>
<tr>
<td>2 Factor Correlated</td>
<td>.944</td>
<td>.909</td>
<td>.873</td>
<td>.073</td>
</tr>
<tr>
<td>3 Factor Uncorrelated</td>
<td>.862</td>
<td>.795</td>
<td>.540</td>
<td>.134</td>
</tr>
<tr>
<td>3 Factor Correlated</td>
<td>.966</td>
<td>.947</td>
<td>.960</td>
<td>.041</td>
</tr>
</tbody>
</table>
Three Correlated-Factor Solution for Subgroups

<table>
<thead>
<tr>
<th>Group</th>
<th>GFI</th>
<th>AGFI</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>.966</td>
<td>.947</td>
<td>.960</td>
<td>.041</td>
</tr>
<tr>
<td>White</td>
<td>.934</td>
<td>.896</td>
<td>.913</td>
<td>.061</td>
</tr>
<tr>
<td>African American</td>
<td>.953</td>
<td>.926</td>
<td>.999</td>
<td>.000</td>
</tr>
</tbody>
</table>
Next Steps

- evaluate internal consistency of items belonging to a single factor (alpha)
- create subscales by averaging or summing scores on indicators belonging to a single factor
- name the factors/subscales
- use the subscales as predictors or outcomes in other analyses
Factor Subscales as Outcomes

Demographics
- race/ethnicity
- age
- SES

Psychosocial
- mental health
- social support
- perceived access

Clinical
- symptoms
- diagnosis

Acute Pain
Factor Analysis References


