Statistics for Survey Design: Solutions for Success

Richard E. Boles, Ph.D.
University of Colorado Denver - Anschutz Medical Campus
Objectives

- Describe the scale purpose
- Identify three core concepts of scale development
- Review scale development and statistical steps
An important term:

*Latent Constructs*: A pattern of activity or concept that cannot be directly measured but has direct or indirect influence on other measured variables.

- The effect of depression on weight
- The effect of parenting styles on child food intake
Survey design: scales

Purpose: to indirectly measure a variable
- Age: single item—no need to further explain!
- Parental feeding styles: measured by multiple items
  - Representing latent, unobservable constructs

Why create yet another scale?
- Scale does not exist or not feasible
- New sample of participants or new theory
  - Different age group, different race/ethnicity
Three core concepts of scale development

Relating theory to measurement
... but not always

Scale Reliability
The level of variable influence on a set of items or... consistency

Scale Validity
Determines if the variable causes item variance or... accuracy
Theoretical based measurement

Constructs: A concept that cannot be directly measured. Why do kids eat certain foods?

- Parental Feeding Styles
- Availability of foods at home
- Parent Race / Ethnicity
- Child Temperament
- Child food intake
I need a measure!

Please look *with exhaustion* in the published literature!

Consider an adaptation of an existing instrument

The original psychometrics are lost, however!
Relating Theoretical Constructs to Measures

- Caregiver Feeding Styles Questionnaire
- Parental Feeding Styles
- Parent Race / Ethnicity
- Child Temperament
- Home Food Inventory
- Availability of foods at home
- Food Frequency Questionnaire
- Child food intake
Atheoretical?

Some scales may develop from descriptive aims that evolve into a more theoretical basis.

- Report on the foods and drinks in your home
- Checklist of items to endorse

While foods in the home sounds descriptive, it can also be linked (statistically) with other behavioral or psychological constructs that affect each other (now part of a theory!)
Scale Reliability
## Reliability

Multiple methods, but first consider response format:

**Continuous versus dichotomous**

1. *I restrict the amount of sweets my child eats*
   - Always: 0
   - Most of the time: 1
   - Sometimes: 2
   - Rarely: 3
   - Never: 4

2. *I currently have apples in the home*
   - Yes: 1
   - No: 0
The Caregiver’s Feeding Style’s Questionnaire:

Comprised of two scales:

- **Demandingness**: The degree to which parents try to get their child to eat
- **Responsiveness**: Child centered strategies (praise and reasoning)
Reliability

Internal consistency of items:
- How well do all of the items seem to measuring a single construct?
- The items correlate well with each other
- Equated with Cronbach’s coefficient alpha: $\alpha$
  (For dichotomous response use Kuder Richardson 20)
So what exactly does computing alpha accomplish?

The variability among items is due to two things:
- True differences in the latent variable across participants (Signal)
- Error. Differences in scores caused by anything else besides the construct (Noise)

Computing alpha separates the signal and noise

Alpha equals the total variance among items that is due to the latent variable (e.g., responsiveness)
Reliability

Improving Alpha improves power!

The ability to detect a statistically significant result (Power) is affected by scale reliability.

In other words, develop a more reliable scale and you reduce measurement error and therefore make it easier to detect statistical differences.
Testing for Alpha: SPSS

1. Click Analyze > Scale > Reliability Analysis

2. You will see the Reliability Analysis dialogue box

3. Transfer items into the box; leave model as “Alpha”

4. Click on Statistics button, select all descriptives and inter-item correlations

Testing for Alpha: SPSS

Interpreting Alpha:

- $\alpha \geq 0.9$  
  Excellent
- $0.8 \leq \alpha < 0.9$  
  Good
- $0.7 \leq \alpha < 0.8$  
  Acceptable
- $0.6 \leq \alpha < 0.7$  
  Questionable
- $0.5 \leq \alpha < 0.6$  
  Poor
- $\alpha < 0.5$  
  Unacceptable
Testing for Alpha: SPSS

<table>
<thead>
<tr>
<th>Item</th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qu1</td>
<td>24.20</td>
<td>45.029</td>
<td>.633</td>
<td>.588</td>
<td>.767</td>
</tr>
<tr>
<td>Qu2</td>
<td>23.93</td>
<td>47.352</td>
<td>.520</td>
<td>.651</td>
<td>.783</td>
</tr>
<tr>
<td>Qu3</td>
<td>24.07</td>
<td>46.629</td>
<td>.654</td>
<td>.899</td>
<td>.767</td>
</tr>
<tr>
<td>Qu4</td>
<td>23.48</td>
<td>47.111</td>
<td>.551</td>
<td>.825</td>
<td>.779</td>
</tr>
<tr>
<td>Qu5</td>
<td>23.60</td>
<td>51.257</td>
<td>.389</td>
<td>.573</td>
<td>.799</td>
</tr>
<tr>
<td>Qu6</td>
<td>24.47</td>
<td>50.695</td>
<td>.72</td>
<td>.693</td>
<td>.702</td>
</tr>
<tr>
<td>Qu7</td>
<td>24.07</td>
<td>45.210</td>
<td>.589</td>
<td>.777</td>
<td>.800</td>
</tr>
<tr>
<td>Qu8</td>
<td>24.20</td>
<td>56.457</td>
<td>.128</td>
<td>.791</td>
<td>.823</td>
</tr>
<tr>
<td>Qu9</td>
<td>24.07</td>
<td>45.210</td>
<td>.589</td>
<td>.610</td>
<td>.774</td>
</tr>
</tbody>
</table>
Question?

When do you decide to drop items from a scale if it appears to improve the overall reliability?

Consider the context of your study:

- New scale? Developing and validating? Yes
- Established scale? Resist dropping as it will make it impossible to compare across studies—at least report both scores (full and reduced scale)—indicate further validation is necessary
Reliability

Other methods of reliability: beyond alpha

Split half correlations:
- First half of items compared to last half
  - Fatigue, failure to finish, practice effects
- Odd items compared to even items
- Random halves

The correlation between the two tests equals the reliability of each set
Reliability

Test-retest reliability

- Same test administered twice over time
- Affected by learning, carry–over, recall
- Generally 2 days to 2 weeks
- May look at longer periods of time and multiple assessments for stability of measure
Validity

Remember to resist making assumptions about scales

Unreliable & Unvalid
Unreliable, But Valid
Reliable, Not Valid
Both Reliable & Valid
Validity

Three main types of scale validity:

- Content validity
- Criterion-related validity
- Construct validity
Content Validity

How well did you sample items to totally reflect the construct?

- Review other measures
- Get expert opinion
- Get participant opinion
Criterion-related Validity

How well does your new scale associate with an identified criterion measure or “Gold Standard”?

The association of a participant’s score on a new test and an established test or an objective measure

- Parent report of available vegetables in the home
- Trained observer recording vegetables in the same home
Construct Validity

How well does your scale assess other similar or dissimilar tests for similar or different constructs?

- **Convergent validity**: your scale is similar to other scales.
- **Divergent validity**: your scale is not related or negatively related to scales of different constructs.

Your new scale on Parental Feeding Responsiveness:

- Parenting Warmth subscale
- Restriction Subscale
Testing validity

Associations and Linear Regressions:

The new test is set as the predictor variable and the criterion variable is the established variable.
Scale development: Some basic steps

What exactly do you want to measure?
- Literature review
- Theoretical development

Start with item development
- Review existing measures
- Qualitative methods: conduct interviews, focus groups, obtain expert opinion and/or participant opinion
Scale development: item pool

Include everything that seems to make sense according to your understanding of the construct; you will eliminate items at a later point.

The goal is to achieve content validity.

How well did you sample items to totally reflect the construct?

Modify existing items or create new items.
Creating new items:

- Make them about one idea. Avoid double barrel questions: I restrict sweets because they make you unhealthy
- Keep the items short and easy to read
- Use reading level assessments and avoid multisyllabic words
- Avoid trendy or colloquial words: Eating cake while on a diet is a FAIL!

Choose a response format:

- Dichotomous (less cognitive labor but limited variability), Likert-type (longer to answer but more variability), Visual Analog Scale (difficult to score but free range of response)
TIME TO FACTOR ANALYZE!
Not Yet...
Check your data for reliable correlations

Sample Size
- Low r reliability from small samples
- 50 = very poor; 100 = poor; 200 = fair; 300 = good; 500 = very good; 1000 = excellent
- 10 observations per variable

Normality
- Skewness: divide the skewness value by the standard error for skewness and what you get is a z score for skewness. Values greater than 3.3 are a problem.
Check your data for reliable correlations

**Outliers**

*Univariate:* very large standardized scores (z scores greater than 3.3) - Analyze -> Descriptives, ... the box at the bottom that says "Save standardized values as variables"

*Multivariate: Mahalanobis Distance test*

Either correct a data entry error, eliminate, reduce the influence, or transform (after checking normality)

**Multicollinearity**

Not an issue for Principal Components Analysis

IS an issue for Factor Analysis

Correlation greater than .90 (drop one)
Are your data appropriate?

Kaiser-Meyer-Olkin (KMO): KMO compares the observed correlation matrix to the partial correlation matrix.

- Cutoffs for the KMO: > .90 = marvelous; in .8 range, meritorious; .7 middling; .6 mediocre; .5 miserable; and < .5, unacceptable.

Bartlett’s Test of Sphericity: This is a test of the null hypothesis that the variables in the population correlation matrix are uncorrelated.

- You want this to be significant.
Structural Tests

- Principal Components Analysis (PCA)
- Factor Analysis (FA)
  - Exploratory Factor Analysis (EFA)
  - Confirmatory Factor Analysis (CFA)
Items for Factor 1

Loaded poorly for Factor 1

Loaded well for both Factors: cross loaded

Items for Factor 2
This is NOT the same as factor analysis!

While the computations and results may look similar, the purposes are indeed different.

PCA provides composites of variables contained from a larger set of variables. The results are based from data and derived from the items. It is considered a data reduction method. It accounts for a portion of the total variance from original items.

FA provides composite variables but they represent hypothetical variables. Factors explain the shared variance.

There are some folks who do not agree with making distinctions between FA and PCA!

(they are wrong)…. 😊
PCA can make the test easier to complete by eliminating items that perform poorly.

Consider this method as a step along psychometric processes but not the final step:

You should not infer underlying constructs from the components.
PCA steps in SPSS

- Click on Analyze, Dimension Reduction, Factor
  - Add all items to variables box

- Click Descriptives:
  - Univariate descriptives
  - Initial solution
  - Correlation Matrix
    - Coefficients
    - Sig. levels
    - KMO/Bartlett’s test
PCA steps in SPSS

Click on Extraction
- Choose Principal Components
- Use the correlation matrix
- Display the \textit{unrotated} factor solution and scree plot
  - Scree plot is graph of eigenvalues
- Extract factors based on eigenvalues greater than 1
  - Eigenvalue is the total variance accounted for by a component
- Save variables using regression method
  - This will allow us to check for correlated components
## Total Variance Explained

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
</tr>
<tr>
<td>1</td>
<td>3.550</td>
<td>23.666</td>
</tr>
<tr>
<td>2</td>
<td>1.294</td>
<td>8.624</td>
</tr>
<tr>
<td>3</td>
<td>1.204</td>
<td>8.026</td>
</tr>
<tr>
<td>4</td>
<td>1.152</td>
<td>7.680</td>
</tr>
<tr>
<td>5</td>
<td>1.055</td>
<td>7.035</td>
</tr>
<tr>
<td>6</td>
<td>.964</td>
<td>6.427</td>
</tr>
<tr>
<td>7</td>
<td>.751</td>
<td>5.004</td>
</tr>
<tr>
<td>8</td>
<td>.702</td>
<td>4.680</td>
</tr>
<tr>
<td>9</td>
<td>.684</td>
<td>4.562</td>
</tr>
<tr>
<td>10</td>
<td>.671</td>
<td>4.476</td>
</tr>
<tr>
<td>11</td>
<td>.651</td>
<td>4.337</td>
</tr>
<tr>
<td>12</td>
<td>.615</td>
<td>4.103</td>
</tr>
<tr>
<td>13</td>
<td>.593</td>
<td>3.956</td>
</tr>
<tr>
<td>14</td>
<td>.587</td>
<td>3.912</td>
</tr>
<tr>
<td>15</td>
<td>.527</td>
<td>3.511</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Check to see if components are related

- Run correlations among the extracted components
  - These new variables were created when you saved variables during initial PCA

- If the components are NOT related use an orthogonal rotation method (e.g., varimax, quartimax, equamax)
  - Rotation methods make the components easier to interpret

- If the components ARE related use an oblique rotation method (e.g., promax, oblimin, direct quartimin)
Re-run it

Now choose the number of components:

- In this case 5 components are to be extracted
- Select the rotation method and examine the rotated component matrix
Review items

- Look for cross loaded items
- Look for small components (fewer than 3 items)
- Primary loadings >= .6 are good (and higher is better)
- Cross-loadings should be small – generally at least .2 less than the primary loading (better if .3 or .4 lower)
  - Items are double-loaded if they have loadings on more than 1 factor that are comparable in size.
- Communalities: the amount of variance for a variable explained by all of the components (good to be above .45)
<table>
<thead>
<tr>
<th>y1</th>
<th>1.682</th>
<th>0.010</th>
<th>0.185</th>
<th>0.089</th>
<th>0.063</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2</td>
<td>0.738</td>
<td>0.181</td>
<td>0.015</td>
<td>0.088</td>
<td>-0.081</td>
</tr>
<tr>
<td>y3</td>
<td>0.725</td>
<td>0.097</td>
<td>0.065</td>
<td>0.104</td>
<td>-0.079</td>
</tr>
<tr>
<td>y4</td>
<td>0.580</td>
<td>0.105</td>
<td>0.185</td>
<td>0.162</td>
<td>0.139</td>
</tr>
<tr>
<td>y5</td>
<td>0.065</td>
<td>0.203</td>
<td>0.031</td>
<td>0.723</td>
<td>0.030</td>
</tr>
<tr>
<td>y6</td>
<td>0.165</td>
<td>0.066</td>
<td>0.168</td>
<td>0.693</td>
<td>-0.142</td>
</tr>
<tr>
<td>y7</td>
<td>0.158</td>
<td>-0.004</td>
<td>0.122</td>
<td>0.709</td>
<td>0.042</td>
</tr>
<tr>
<td>y8</td>
<td>0.145</td>
<td>0.043</td>
<td>0.752</td>
<td>0.032</td>
<td>-0.009</td>
</tr>
<tr>
<td>y9</td>
<td>0.083</td>
<td>0.220</td>
<td>0.721</td>
<td>0.107</td>
<td>0.021</td>
</tr>
<tr>
<td>y10</td>
<td>0.166</td>
<td>0.136</td>
<td>0.644</td>
<td>0.214</td>
<td>-0.019</td>
</tr>
<tr>
<td>y11</td>
<td>0.082</td>
<td>0.744</td>
<td>0.071</td>
<td>0.139</td>
<td>0.041</td>
</tr>
<tr>
<td>y12</td>
<td>0.200</td>
<td>0.713</td>
<td>0.126</td>
<td>-0.042</td>
<td>-0.103</td>
</tr>
<tr>
<td>y13</td>
<td>0.067</td>
<td>0.712</td>
<td>0.174</td>
<td>0.158</td>
<td>0.063</td>
</tr>
<tr>
<td>y14</td>
<td>0.021</td>
<td>0.053</td>
<td>-0.183</td>
<td>0.073</td>
<td>0.717</td>
</tr>
<tr>
<td>y15</td>
<td>-0.003</td>
<td>-0.046</td>
<td>0.177</td>
<td>-0.119</td>
<td>0.690</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.
### Total Variance Explained

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
<th>Rotation Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
<td>Cumulative %</td>
</tr>
<tr>
<td>1</td>
<td>3.549</td>
<td>27.303</td>
<td>27.303</td>
</tr>
<tr>
<td>2</td>
<td>1.292</td>
<td>9.936</td>
<td>37.239</td>
</tr>
<tr>
<td>3</td>
<td>1.192</td>
<td>9.170</td>
<td>46.409</td>
</tr>
<tr>
<td>4</td>
<td>1.139</td>
<td>8.764</td>
<td>55.173</td>
</tr>
<tr>
<td>5</td>
<td>.757</td>
<td>5.823</td>
<td>60.996</td>
</tr>
<tr>
<td>6</td>
<td>.715</td>
<td>5.501</td>
<td>66.497</td>
</tr>
<tr>
<td>7</td>
<td>.688</td>
<td>5.292</td>
<td>71.789</td>
</tr>
<tr>
<td>8</td>
<td>.675</td>
<td>5.190</td>
<td>76.979</td>
</tr>
<tr>
<td>9</td>
<td>.655</td>
<td>5.038</td>
<td>82.017</td>
</tr>
<tr>
<td>10</td>
<td>.623</td>
<td>4.789</td>
<td>86.806</td>
</tr>
<tr>
<td>11</td>
<td>.597</td>
<td>4.591</td>
<td>91.397</td>
</tr>
<tr>
<td>12</td>
<td>.592</td>
<td>4.550</td>
<td>95.947</td>
</tr>
<tr>
<td>13</td>
<td>.527</td>
<td>4.053</td>
<td>100.000</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
### Rotated Component Matrix

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y1</td>
<td>0.680</td>
<td>0.006</td>
<td>0.200</td>
<td>0.080</td>
</tr>
<tr>
<td>y2</td>
<td>0.738</td>
<td>0.180</td>
<td>0.016</td>
<td>0.092</td>
</tr>
<tr>
<td>y3</td>
<td>0.725</td>
<td>0.094</td>
<td>0.070</td>
<td>0.107</td>
</tr>
<tr>
<td>y4</td>
<td>0.581</td>
<td>0.118</td>
<td>0.161</td>
<td>0.171</td>
</tr>
<tr>
<td>y5</td>
<td>0.064</td>
<td>0.205</td>
<td>0.023</td>
<td>0.727</td>
</tr>
<tr>
<td>y6</td>
<td>0.165</td>
<td>0.067</td>
<td>0.157</td>
<td>0.704</td>
</tr>
<tr>
<td>y7</td>
<td>0.156</td>
<td>-0.003</td>
<td>0.118</td>
<td>0.711</td>
</tr>
<tr>
<td>y8</td>
<td>0.143</td>
<td>0.039</td>
<td>0.765</td>
<td>0.025</td>
</tr>
<tr>
<td>y9</td>
<td>0.081</td>
<td>0.216</td>
<td>0.731</td>
<td>0.101</td>
</tr>
<tr>
<td>y10</td>
<td>0.164</td>
<td>0.128</td>
<td>0.666</td>
<td>0.202</td>
</tr>
<tr>
<td>y11</td>
<td>0.081</td>
<td>0.743</td>
<td>0.073</td>
<td>0.137</td>
</tr>
<tr>
<td>y12</td>
<td>0.201</td>
<td>0.717</td>
<td>0.108</td>
<td>-0.026</td>
</tr>
<tr>
<td>y13</td>
<td>0.066</td>
<td>0.711</td>
<td>0.179</td>
<td>0.154</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.
EFA and CFA

- EFA can identify a testable model (no assumptions about which items load with which latent factor)
- CFA defines a model a priori and tests for support (you define which items are related to specific constructs (latent variables))
Identifies the number and nature of latent variables based on the observations:

That is, what caused the observed responses?

OFTEN CONFUSED WITH PCA!

Looks very similar to PCA in SPSS except the method you use will be “Maximum Likelihood or Principal Axis”

PCA: uses the total variance from items (common and unique) does not assume error variance

EFA: uses only the common variance from items
CFA

Much more complex than PCA and EFA

Tests whether or not a specific set of constructs (latent variables) influence the responses in a predicted way

Does the model “fit”?

Assessment of goodness of fit statistics

The model is “plausible”

Often requires specialized software

LISREL, EQS, AMOS, Mplus, SAS
Survey development and validation is a...  

... process and not an event!

Pay close attention to the published literature to avoid reinventing the wheel

Focus on developing a theoretical basis-when possible

Establish reliability and validity

PCA to reduce items

EFA to identify a model with latent constructs

CFA to test the plausibility of that model
Thank you!