Constructing Indices and Scales

Hsueh-Sheng Wu

CFDR Workshop Series

June 8, 2015

BGSU
Center for Family and Demographic Research

Outline

- What are scales and indices?
- Graphical presentation of relations between items and constructs for scales and indices
- Why do sociologists need scales and indices?
- Similarities and differences between scales and indices
- Constructions of scales and indices
- Criteria for evaluating a composite measure
- Evaluation of scales and indices
- How to obtain the sum score of a scale or an index?
- Conclusion

What Are Scales and Indices?

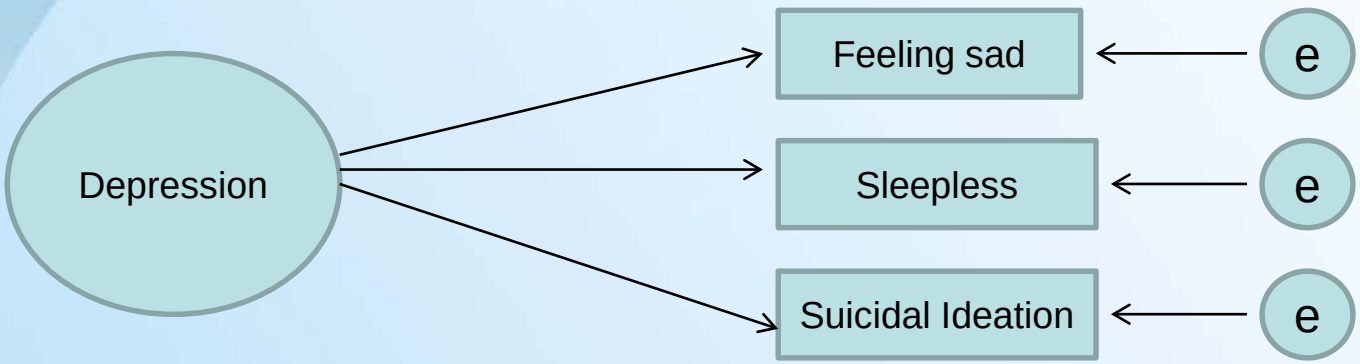
- Scales and indices are composite measures that use multiple items to collect information about a construct. These items are then used to rank individuals.

- Examples of scales:
  - Depression scales
  - Anxiety scale
  - Mastery scale

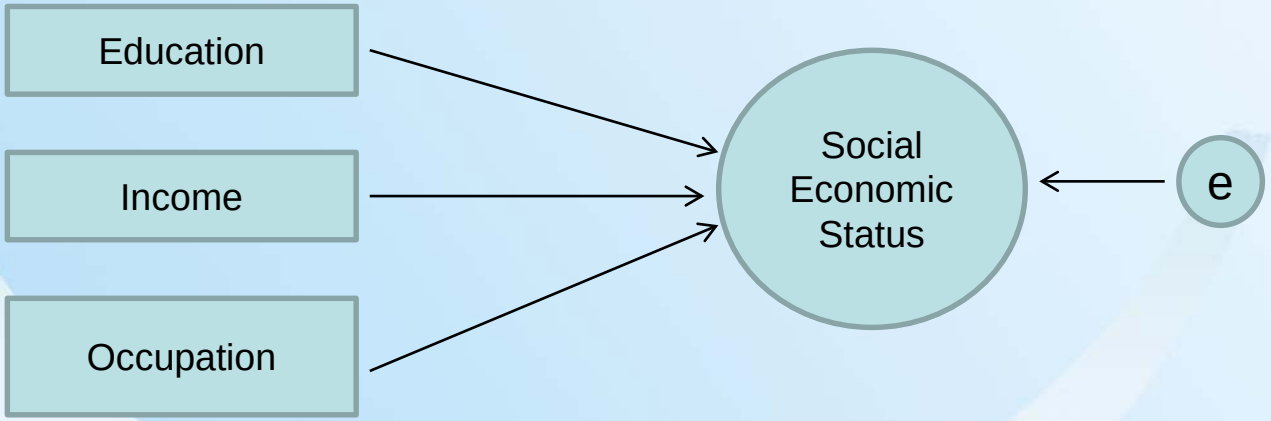
- Examples of indices:
  - Socio-Economic Status (SES) index
  - Consumer price index
  - Stock market index
  - Body mass Index

Graphical Presentation of Relations between Items and Constructs for Scales and Indices

- Scale:



- Index:



Why Do Sociologists Need Scales and Indices?

- Most social phenomenon of interest are multi-dimensional constructs and cannot be measured by a single question, for example:
  - Well-being
  - Violence
- When a single question is used, the information may not be very reliable because people may have different responses to a particular word or idea in the question.
- The variation of one question may not be enough to differentiate individuals.
- Scales and indices allow researchers to focus on large theoretical constructs rather than individual empirical indicator.

Similarities and Differences between Scales and Indices

Similarities:

- Both try to measure a composite construct or constructs.
- Both recognize that the construct or constructs have multiple-dimensional attributes.
- Both use multiple items to capture these attributes.
- Both can apply various measurement levels (i.e., nominal, ordinal, interval, and ratio) to the items.
- Both are composite measures as they both aggregate the information from multiple items.
- Both use the weighted sum of the items to assign a score to individuals.
- The score that an individual has on an index or a scale indicates his/her position relative to those of other people.

Differences:

- Scale consists of effect indicator, but index includes causal indicators
- Scales are always used to give scores at individual level. However, indices could be used to give scores at both individual and aggregate levels.
- They differ in how the items are aggregated.
- Many discussions on the reliability and validity of the scales, but few discussions on those of indices.

Construction of Scales

- DeVellis, Robert (2011) Scale Development: Theory and Applications

1. Determine clearly what it is you want to measure
2. Generate an item pool
3. Determine the format for measurement
4. Have the initial item pool reviewed by experts
5. Consider inclusion of validation items
6. Administer items to a development sample
7. Evaluate the items
8. Optimize scale length

Construction of Indices

Babbie, E (2010) suggested the following steps of constructing an index

1. Selecting possible items
   - Decide how general or specific your variable will be
   - Select items with high face validity
   - Choose items that measure one dimension of the construct
   - Consider the amount of variance that each item provides

2. Examining their empirical relations
   - Examine the empirical relations among the items you wish to include in the index

3. Scoring the index
   - you then assign scores for particular responses, thereby making a composite variable out of your several items

4. Validating it
   - item analysis
   - The association between this index and other related measures.

Concepts of Reliability and Validity

- Reliability: Whether a particular technique, applied repeatedly to the same object, yields the same result each time
  - Test-retest reliability
  - Alternate-forms reliability (split-halves reliability)
  - Inter-observer reliability
  - Inter-item reliability (internal consistency)
- Validity: The extent to which an empirical measure adequately reflects the real meaning of the concept under consideration
  - Face validity
  - Content validity
  - Construct validity (convergent validity and discriminant validity)
  - Criterion validity (concurrent validity and predictive validity)

Reliability

- Test-retest reliability:
  - Apply the test at two different time points. The degree to which the two measurements are related to each other is called test-retest reliability
  - Example: Take a test of math ability and then retake the same test two months later. If receiving a similar score both times, the reliability of this test is high.
  - Possible problems:
    - Test-retest reliability holds only when the phenomenon do not change between two points in time.
    - Respondents may get a better score when taking the same test the second time, which reduce the test-retest reliability

Reliability (cont.)

- Alternate-forms reliability:
  - Compare respondents' answers to slightly different versions of survey questions. The degree to which the two measurements are related to each other is called alternative-form reliability.
  - Possible problem: How to make sure these two alternate-forms are equivalent?

Reliability (cont.)

- Split-halves reliability
  - Similar to the concept of alternate-forms reliability
  - Randomly divide survey sample into two. These two halves of the sample answer two forms of the questions. If the responses of the two halves of the sample are about the same, the measure's reliability is established
  - Possible problem: What if these two halves are not equivalent?

Reliability (cont.)

- Inter-observer reliability
  - When more than one observer to rate the same people, events, or places
  - If observers are using the same instrument to rate the same thing, their ratings should be very similar. If they are similar, we can have much confidence that the ratings reflect the phenomenon being assessed than the orientations of the observers
  - Possible problem: The reliability is established for observers, not for the measurement items. Thus, inter-observer reliability cannot be generalized to studies with different observers.

Reliability (cont.)

- Inter-item reliability
  - Apply only when you have **multiple items** to measure a single concept.
  - The stronger the association among the individual items, the higher the reliability of the measures.
  - In Statistics, we use Cronbach's Alpha to measure inter-item reliability.
- Possible problems
  - You can increase the value of Cronbach's Alpha by increasing the number of scale items even if these items are not highly correlated.

Validity

- The extent to which an empirical measure adequately reflects **the real meaning of the concept** under consideration
  - Face validity
  - Content validity
  - Criterion validity (concurrent and predictive validity)
  - Construct validity (convergent and discriminant validity)

Face Validity

- The quality of an indicator that makes it seem a reasonable measure of some variable ("on its face")

  Example: frequency of church attendance – an indicator of a person's religiosity

Content Validity

- The degree to which a measure covers **the full range** of the concept's meaning

  Example: Attitudes toward police department contain different domains, for example, expectation, past experience, others' experience, and mass media

Criterion Validity

- The degree to which a measure relates to some **external criterion**

  Example: People who score high on a depression scales also are more likely to receive a diagnosis of clinical depression.

Concurrent Validity

- A measure yields scores that are closely related to scores obtained by using previously established measures of the same construct.

  Example: Comparing a test of sales ability to the person's sales performance

Predictive Validity

- The ability of a measure to predict score on a criterion measured **in the future**

  Example: SAT scores can predict the college students' GPA (SAT scores would be a valid indicator of a college student's success)

Construct Validity

- The degree to which a measure **relates to other variables** as expected within a system of theoretical relationships

  Example: If we believe that marital satisfaction is related to marital fidelity, the response to the measure of marital satisfaction and the response to the measure of marital fidelity should act in the expected direction (i.e., more satisfied couples also less likely to cheat each other)

Convergent vs. Discriminant Validity

- Convergent validity -- one measure of a concept is associated with different types of measures of **the same concept**

- Discriminant validity -- one measure of a concept is **not** associated with measures of **different** concepts

Reliability and Validity of Scales and Indices

Table 1. The reliability and validity of scales and indices

	Scales	Indices
Reliability		
Test-retest reliability	X	X
Alternate-forms reliability (split-halves reliability)	X	?
Inter-observer reliability	X	X
Inter-item reliability	X	?
Validity		
Face validity	X	X
Content validity	X	X
Criterion validity		
concurrent validity	X	X
predictive validity	X	X
Construct validity		
convergent validity	X	X
discriminant validity	X	X

How to obtain the sum score of a scale or an index

- Common way
  - Assume that each item have the equal weight, and simply sum each item together
- Use factor analysis

Table 1. Variable about environment in General Socail Survey, 2010

Variable Name	Variable Description
grncon	concerned about environment
grndemo	protested for envir issue
grnecon	worry too much about envir, too little econ
grneffme	environment effect everyday life
grnexagg	environmental threats exaggerated

How to obtain the sum score of a scale or an index (Cont.)

```
. factor grncon grndemo grnecon grneffme grnexagg
(obs=1287)

Factor analysis/correlation                       Number of obs    =     1287
    Method: principal factors                     Retained factors =        3
    Rotation: (unrotated)                         Number of params =       10
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.14876	1.09574	1.3943	1.3943
Factor2	0.05302	0.03888	0.0643	1.4586
Factor3	0.01414	0.18208	0.0172	1.4758
Factor4	-0.16794	0.05614	-0.2038	1.2720
Factor5	-0.22408	.	-0.2720	1.0000

```
    LR test: independent vs. saturated:  chi2(10) =  670.67 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances
```

Variable	Factor1	Factor2	Factor3	Uniqueness
grncon	0.5491	-0.0853	0.0352	0.6900
grndemo	-0.1385	0.0054	0.1125	0.9682
grnecon	0.5622	0.1020	-0.0046	0.6735
grneffme	-0.3677	0.1678	0.0138	0.8365
grnexagg	0.6139	0.0846	0.0064	0.6160

BGSU Center for Family and Demographic Research

How to obtain the sum score of a scale or an index (Cont.)

Factor1 = 0.54910*grncon + -0.1385*grndemo + 0.5622*grnecon + -0.3677*grneffme + 0.6139*grnexagg
Factor2 = -0.0853*grncon + 0.0054*grndemo + 0.1020*grnecon + 0.16780*grneffme + 0.0846*grnexagg
Factor3 = 0.03520*grncon + 0.1125*grndemo + -0.0046*grnecon + 0.01380*grneffme + 0.0064*grnexagg

How to obtain the sum score of a scale or an index (Cont.)

Using principal component analysis for indices

```
. pca  educ realrinc prestg80

Principal components/correlation                  Number of obs    =      1200
                                                  Number of comp.  =         3
                                                  Trace            =         3
    Rotation: (unrotated = principal)             Rho              =    1.0000
```

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	1.8637	1.21591	0.6212	0.6212
Comp2	.647785	.159268	0.2159	0.8372
Comp3	.488517	.	0.1628	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Unexplained
educ	0.5964	-0.3750	-0.7097	0
realrinc	0.5382	0.8428	0.0070	0
prestg80	0.5955	-0.3862	0.7044	0

Score1 = 0.5964 *educ + 0.5382*realrinc + 0.5955*prestg80
Score2 = -0.3750 *educ + 0.8428*realrinc + -0.3862*prestg81
Score3 = -0.7097 *educ + 0.0070*realrinc + 0.7044*prestg82

Conclusion

- Both indices and scales are composite measures that use multiple items to collect information about a construct.
- With indices and scales, researchers can move beyond examining observable indicators, and start examining abstract construct.
- Scales have effect indicators, while indices have causal indicators.
- To determine whether your are constructing an index or a scale, you need to think about whether these indicators are highly correlated with each other.
- If you have any questions about measuring a construct or constructs, please come see me at 5D, Williams Hall or send me an email (wuh@bgsu.edu).