Session 6: Assessing Measurement Validity

Are we measuring what we think we are measuring?

Dr. Nicholas A. Valentino

Remember!

- A measure is composed of truth and error
- Observed score, true score, measurement error
  - $X = t + e$
    - X is the observed score
    - t is the true score
    - e is the error

We must distinguish random error (affects variation, not the average score) from systematic error (bias affects the average score)

$$\text{Observed Score} = \text{True Score} + \left(\text{Random Error} + \text{Systematic Error} \right)$$

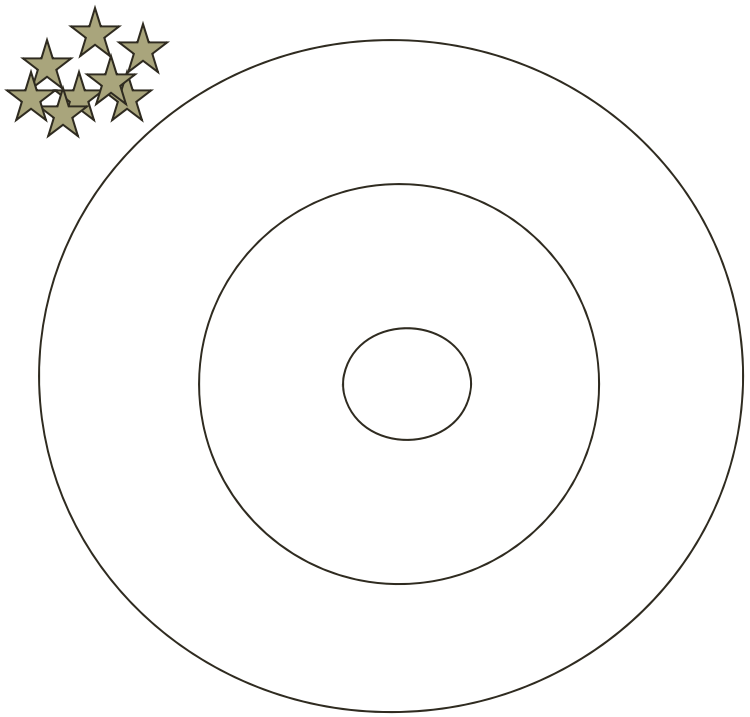
RELIABILITY AND VALIDITY: MINIMZING BOTH TYPES OF MEASUREMENT ERRORS

- **Reliability**: How consistent is your measure?
  - Maximize by reducing random error
- **Validity:** Are you measuring what you think you are?
  - Maximize by reducing bias.

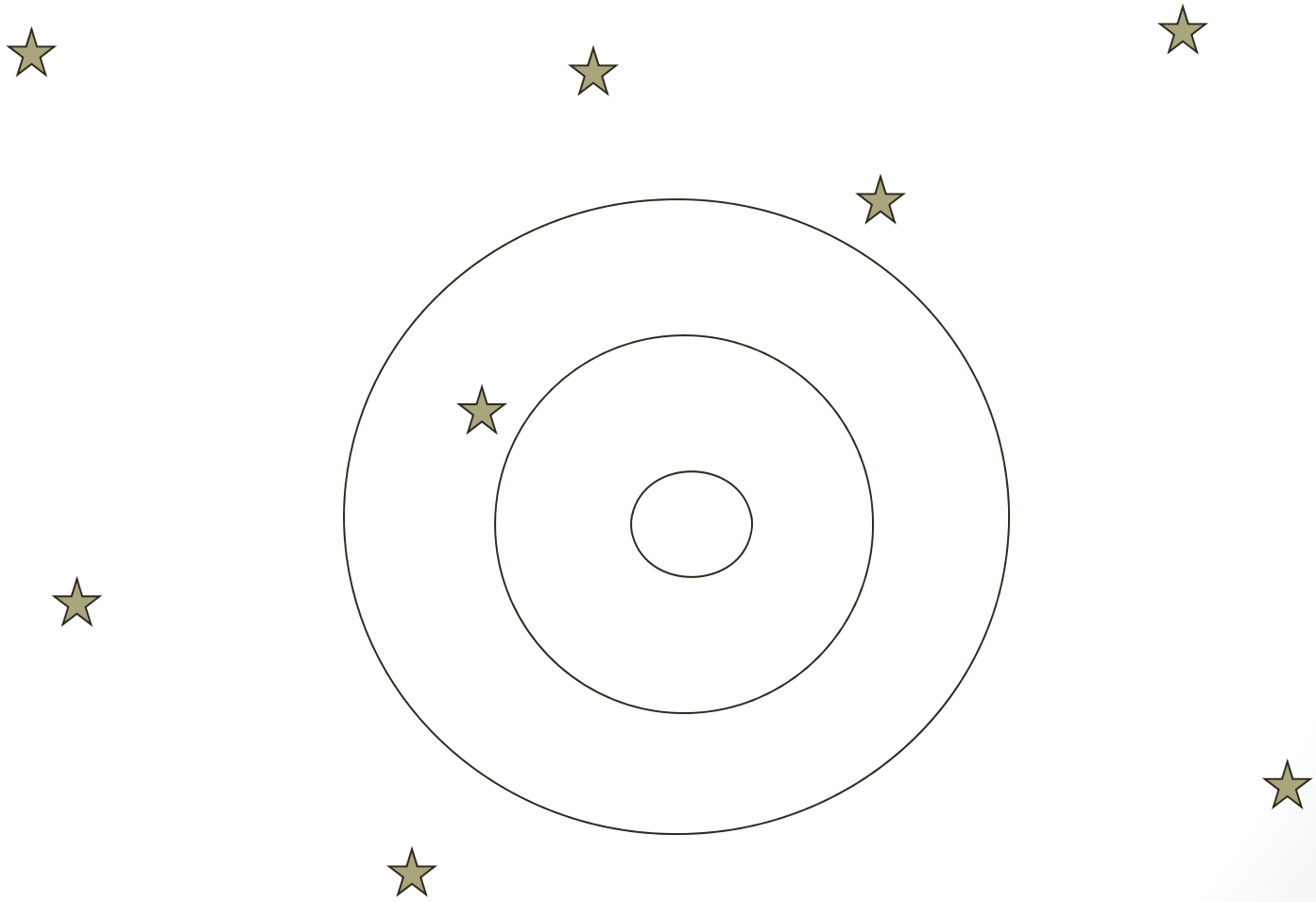
Think of a concept as a target

- Your indicators are arrows.
- We shoot our arrows at the target, and try to hit the bull's-eye every time.

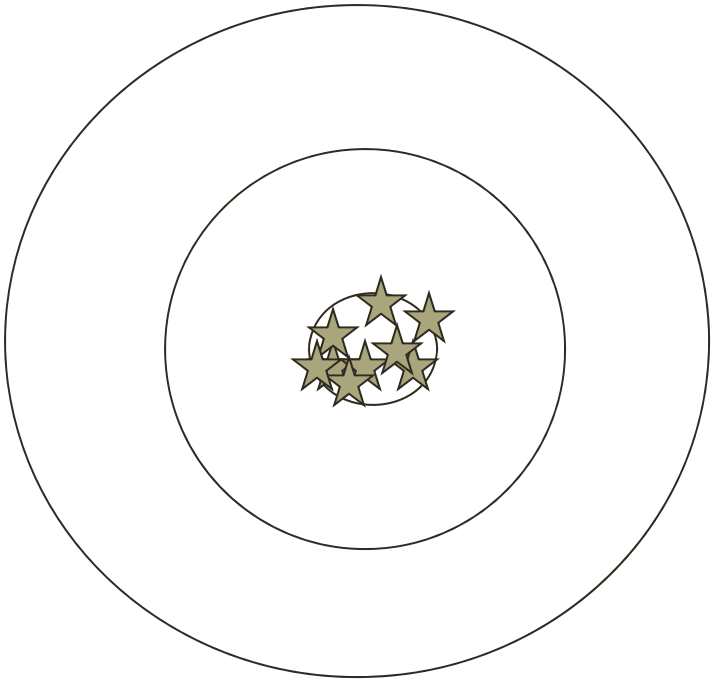
Reliable, (But Not Valid)



Not Valid, Not Reliable



Valid and Reliable



Validity: Does our indicator accurately reflect the concept it is intended to measure?

- First we evaluate without statistics
  - Face Validity: Is it reasonable to think this is a good indicator?
    - On its face, would knowing what kind of car someone drives be a good indicator of their wealth?
    - What if someone gave them the car? What if they went in debt to buy the car?

Validity: Does our indicator accurately reflect the concept it is intended to measure?

- First we evaluate without statistics
  - Content Validity: Does the indicator *cover* the entire concept?
    - Would knowing *only* whether a country censored the news media be enough to measure if it was a “democracy”?
    - We might also want to know if it held open elections, protected citizens’ freedom of speech, guaranteed due process under the law.

Validity: Does our indicator accurately reflect the concept it is intended to measure?

- We could also evaluate validity statistically
  - Criterion-Related Validity: Does the indicator predict some external criterion?
    - Survey indicator of aggression (X_1): "Once in a while I cannot control my urge to strike another person"
    - External Criterion (X_2): Red cards received per season in football.
    - Test for criterion validity with $r(X_1, X_2)$.

Validity: Does our indicator accurately reflect the concept it is intended to measure?

- We could also evaluate validity statistically
  - Construct Validity: Is the indicator correlated with theoretically related constructs?
    - Comes in *convergent* and *discriminant* types
  - Previous theory: self-esteem leads to life satisfaction.
  - Indicator of self esteem (X): "I am able to do things as well as most other people."
  - Indicator of life satisfaction (Y): "In most ways my life is close to the ideal"
  - Test for convergent validity with r(X, Y).

Example of Criterion Validity Test: Are self-reports always biased?

- Do people under-report smoking due to social desirability pressure? (Yeager & Krosnick, 2010).
  - Survey indicator of smoking behavior:
    - "Do you now smoke cigarettes some days, most days, or not at all?"
  - External criterion: Blood samples were analyzed for levels of serum cotinine, an indicator of nicotine exposure.

RESULTS

- On average, only 1.17% to 1.25% of adult respondents who said that they did not use a product containing nicotine nonetheless had high cotinine levels.
- In other words, almost no one lied.
- Lesson: Do not assume self-reports are biased, do a validity check!

Thought exercise

- Using the handouts, write down some thoughts about the validity of the various measures listed.
- Take a few minutes, and then we will discuss your answers as a group.

Problems?

- Evaluate the *face* validity of the following survey measures. Are these questions reasonable for measuring the underlying concept? Why or why not? Write some notes below each question and then we can talk about each one.

- Concept: Anxiety
- Measure: “In general, how often does your heart start beating fast when you think about things you are going to do in the future?”

- Very often
- Somewhat often
- Not too often
- Occasionally
- Not at all

- **What other reasons might people have for a fast heartbeat? Could they be excited and happy, not just anxious? What about angry?**

Problems?

- Concept: Political Knowledge
- Measure: “How many countries currently belong to the Arab League of States?”

- **Is this a good indicator of intelligence, or is it a very specific type of knowledge that many intelligent people will not know?**

Problems?

- Evaluate the *content* validity of the following items. Discuss aspects you would add to measure other dimensions of the construct.

- Concept: Satisfaction with the Museum of Islamic Art
- Measure: On your last visit to the Museum of Islamic Art, how satisfied were you with the
  - Parking facilities
  - Contents of the displays

Other dimensions to include?

- **What about the staff?**
- **What about the guides/maps?**
- **Facilities for the children?**
- **Cleanliness of the facilities?**
- **Clarity of the signs?**

Evaluate the *criterion* validity of the following indicator. Discuss which of these external criteria would be best for evaluating the validity of a survey indicator of safe driving ability, and explain why.

- Concept: Driving ability
- Measures:
- Survey measure: Test of knowledge of rules of the road. "What is the speed limit on the highway?" "How many car lengths between cars is safe for highway driving" "After you have entered a roundabout, do you yield to people entering on the right?"
-
- External criteria: 1. Finding shortest route in trip from one location to another. 2. Number and severity of traffic violations in previous 12 months. 3. How quickly can one find a parking spot in busy parking lot.

-
- **1. Would knowing the shortest route to get from one place to another really be a good indicator of whether you are a safe driver?**
-
- **2. Traffic violations might be a good external criterion for safe driving.**
-
- **3. Being good at finding a parking spot is not necessarily related to how safe you would be on the road.**

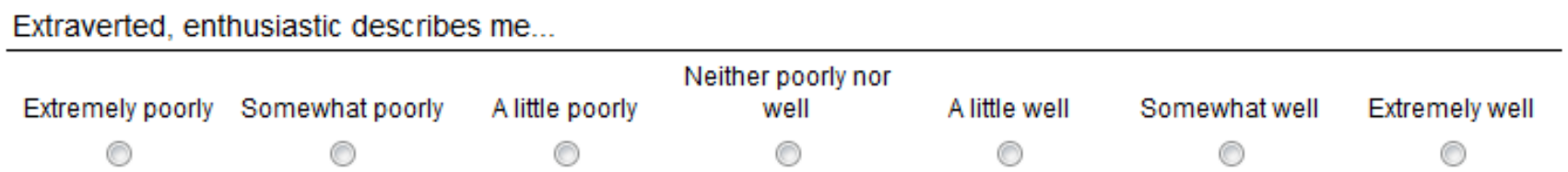
Computer Exercise

- We want to validate our indicator of personality dimensions.
  - Let's try two tests of *construct validity*: Convergent and discriminant
  - Previous theory suggests extraversion should be correlated with talkativeness.
  - Anxiety should NOT be correlated with talkativeness.

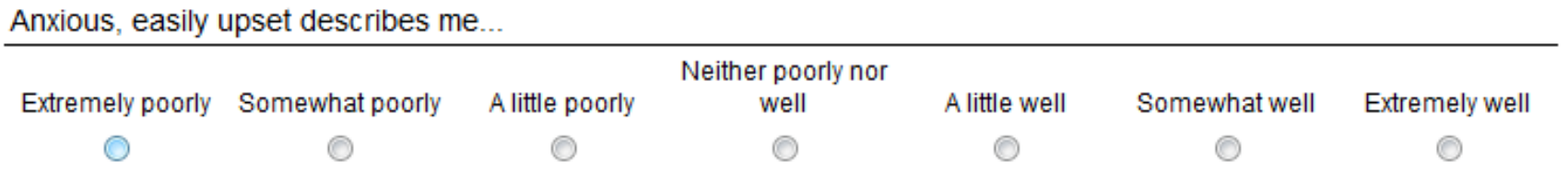
Measuring Personality dimensions

- "We're interested in how you see yourself. Please mark how well the following pair of words describes you, even if one word describes you better than the other."

Extraversion



Anxiousness



Measure of Talkativeness

- “Next we would like you to take a few minutes to describe your life in Qatar. Please tell us what you enjoy most about living here. Also, feel free to mention any aspects of life here you wish were different.”
- We need a simple word count as an indicator of talkativeness

Sample of Responses

- “I enjoy living in Qatar, primarily because it is a calm place, and people are so kind.”
- “I like that it is not too overwhelming, slow paced and easy access to anywhere you want to travel to.”
- “The thing I enjoy most about life in Qatar is multiculturalism. Here in Doha , you can meet people from all walks of life, which helps for understanding other cultures very well.”
- “Social life.”

Correlation

- A correlation is a measure of the degree to which two items “travel together”
- The “Pearson’s r” runs from -1 to 1
- We can calculate the correlation between two variables using the *Analyze / Correlate / Bivariate* menu option.

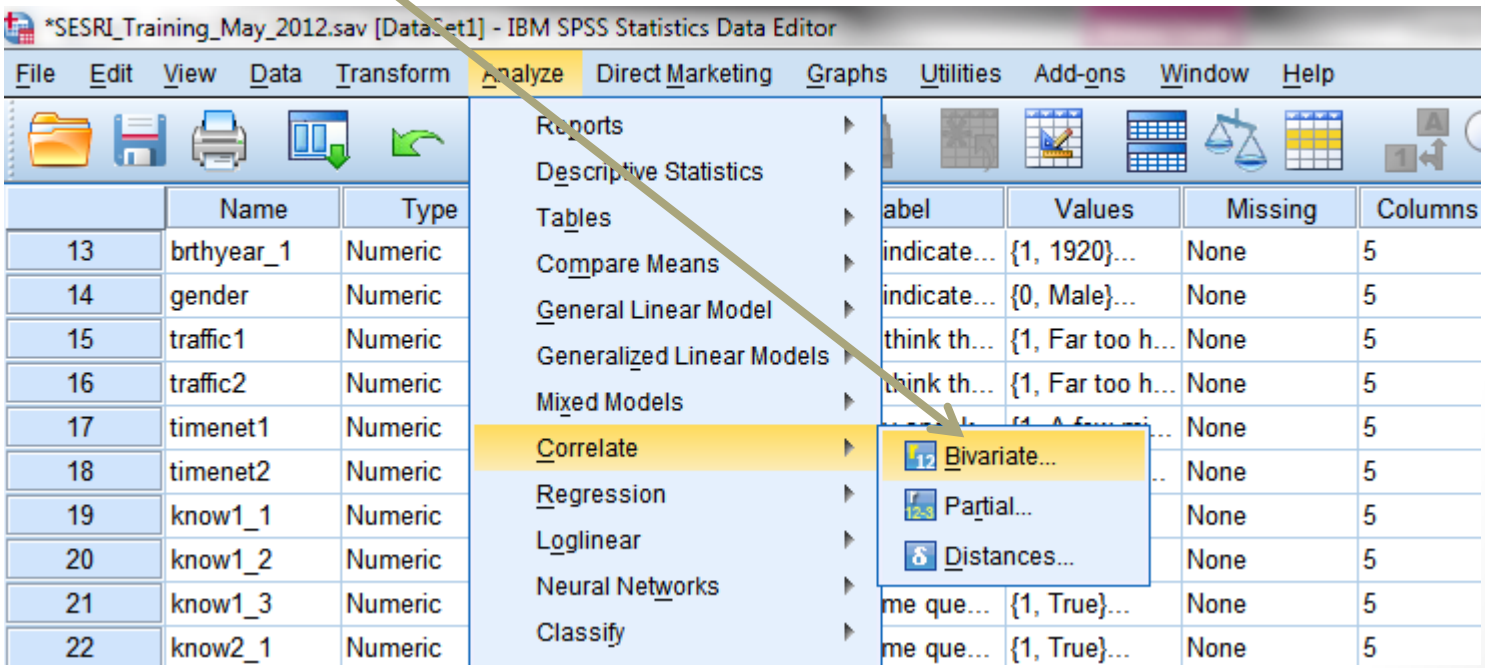
Computer Exercise

- Correlate each of two items (anxious & extravert) measuring two personality dimensions with a variable indicating the number of words you typed in response to question about what you most enjoy about living in Qatar.

- Test our *measurement* hypotheses:
  - 1. Extraversion will be positively correlated with talkativeness.
  - 2. Anxiety will NOT be correlated with talkativeness.

Producing the Correlations Between Personality Dimensions and Talkativeness

Select Menu Option: "Correlate" and "Bivariate"

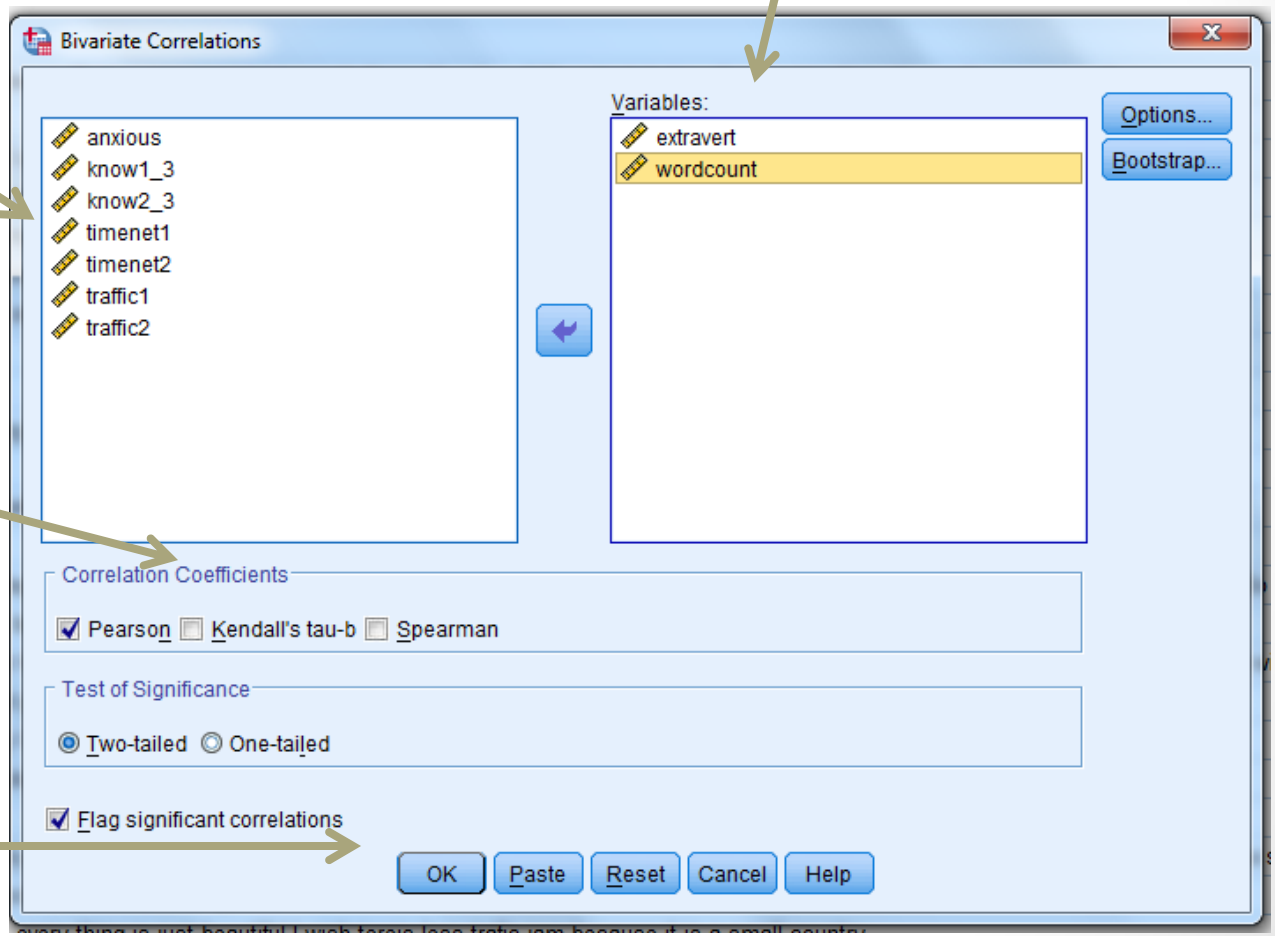


Correlations in SPSS

Select variables from the list on the left. Press the arrow button to move them to the box on the right until you have indicated all variables for which you want pair-wise correlations.

Select Pearson correlation coefficient for interval variables.

Click the OK button.



Correlations

		extravert Extraverted, enthusiastic describes me...	wordcount Number of words R wrote about life in Qatar
extravert Extraverted, enthusiastic describes me...	Pearson Correlation	1	.341
	Sig. (2-tailed)		.111
	N	26	23
wordcount Number of words R wrote about life in Qatar	Pearson Correlation	.341	1
	Sig. (2-tailed)	.111	
	N	23	24

		anxious Anxious, easily upset describes me...	wordcount Number of words R wrote about life in Qatar
anxious Anxious, easily upset describes me...	Pearson Correlation	1	.008
	Sig. (2-tailed)		.972
	N	27	24
wordcount Number of words R wrote about life in Qatar	Pearson Correlation	.008	1
	Sig. (2-tailed)	.972	
	N	24	24

Bibliography

- Bushman, B. J., & Wells, G. L. (1998). Trait aggressiveness and hockey penalties: Predicting hot tempers on the ice. Journal of Applied Psychology, 83, 969-974.
- Yeager, D. S., & Krosnick, J.A. (2010). The validity of self-reported nicotine product use in the 2001-2008 National Health and Nutrition Examination Survey. Medical Care 48(12) 1128 - 1132.
- Hoyle, Rick H., Monica J. Harris, and Charles M. Judd. 2002. Research Methods in Social Relations, 7th edition. Wadsworth.
- Sullivan, John, and Stanley Feldman. 1979. Multiple Indicators. Newbury Park: Sage.

Thank you!