

## Measurement in Social Research—Surveys and Interviews

*“England and America are two countries divided by a common language.”*  
George Bernard Shaw

- I. Feedback about Written Assignment #1
- II. APA Style and Writing the Literature Review—
  - A. Read the “APA Style Sheet” on the course web page
  - B. Citation and bibliography styles
- III. Study questions for Week 5
  1. What’s the difference between “validity” and “reliability”? Give an example of each.
  2. In the article “The cultural adaptation of the Youth Quality of Life Instrument” the authors use the term “cultural equivalence.” Define it, and describe a few of the components.
  3. In the same article, researchers recommended changes not only to the Spanish version of the instrument, but also the English version. Why?
  4. Describe the difference between open-ended and closed-ended questions. Give examples of each.

### IV. Handling Measurement Error--Validity and Reliability

The presence of error is not an “all or nothing” thing. It has to be measured. How do we measure error so we can avoid it? There are two types of ways to measure error, roughly corresponding to random and systematic error: reliability and validity. There are various ways to test reliability and validity. *Not all are required for each measure or instrument.*

**\*\*\*Student report: What’s the difference between “validity” and “reliability”?  
Give an example of each.**

- A. **Reliability (of measure)—concerned with *consistency of measurement*, and it addresses *random error*.** Reliability is whether or not a particular technique done repeatedly would yield the same result each time. There are four main indicators of

reliability:

1. **Interrater reliability**—two researchers, same ratings
2. **Test-retest reliability**—same response from first test to re-test
3. **Alternate forms reliability**—same response from slightly different versions of same measure
4. **Internal consistency reliability** & use of coefficient alpha—similar questions (about the same concept) in survey are answered in the same way.

*\*\*\*Look for coefficient alphas in your literature review. It's a number that tells you how well the questions in an instrument correlate with each other when the instrument is tested (you want a high number). You will need to report this for any instruments you use in your study. A coefficient alpha is considered good if it is 80 or above. If it is not, then you might still use the instrument, but explain the limitation that it has questionable reliability.*

**B. Validity—concerned with accuracy of meaning; addresses systematic error**

1. **Face validity**—does measure reflect most experts' understanding of concept?
2. **Content validity**—does measure reflect range of meaning in the concept? Example: “self esteem” can be many things (dimensions)—observable behaviors, opinions about oneself, perceptions about how other people see you. Do the survey or instrument questions cover these dimensions?
3. **Criterion-related validity**—does a measure of concept agree with another external valid or clearcut concept (or criterion)?
  - a) **Known-groups validity** (most typical). Can a measure differentiate between two or more groups who are already known (in other ways) to have a particular characteristic? Example: Can an instrument measuring depression differentiate those who were diagnosed independently by a clinician with depression from those who were assessed as not being depressed? (The criterion is “clinical diagnosis of depression.”)
  - b) **Predictive validity**—Can a measure of early signs of depression predict that later a formal clinical diagnosis of depression would occur? (Note: this type of validity often applies to instruments of risk assessment, such as future risk of child abuse)
  - c) **Concurrent validity**— Can a measure of current depression accurately match a clinician's diagnosis of depression?
4. **Construct validity**—how measure relates to measures of other concepts (constructs) or variables linked theoretically (used when there are no validated

criteria existing or accessible). Example: since domestic violence and low marital satisfaction are known to be related, your new set of questions on marital satisfaction correlates well with a set of questions on predicting domestic violence. Another way to look at construct validity: an hypothesis says, according to theory, A causes B. If there is a positive relationship (correlation) between the A item responses and the B item responses on your instrument, there is high construct validity.

\*\*\*What is a “construct”?

- a) **Convergent validity**—results of measure correspond to results of other measure(s) measuring a related construct (such as a depression scale score corresponding to a measure of suicidal thoughts)
- b) **Discriminant validity**—results of measure are correlated with other similar measure, but not correlated with items or measure of something different (such as a depression scale score corresponding to thoughts of suicide, but not corresponding to symptoms of hyperactivity or grandiosity)

\*\*Name the type of validity or reliability implied in these scenarios:

1. Validity studies show that a set of questions about alcohol dependence correlates well with the observed daily amount of alcohol consumption.
2. Responses to questions using the Rosenberg Self Esteem Scale correlate highly with each other consistently from test to re-test
3. Most experts would agree that the signs of Attention Deficit Disorder with Hyperactivity include (a) a child’s inability to focus attention at school and home, (b) difficulty staying on tasks (even those the child likes) and (c) impulsivity.
4. Clients who have been identified as having low levels of marital satisfaction on your new written set of questions also test as having low marital satisfaction on another instrument measuring the same thing.
5. Two clinical observers of parent-child interaction mostly agree on whether the parent’s response to the child’s behavior is flexible or rigid
5. A researcher wants to make sure that the instrument she is developing covers all the various dimensions of the concept “empowerment.”
6. A new instrument called the Self Perceived Empowerment Scale has an alpha coefficient of .85

\*\*\*Note: You can have good Reliability *without* good validity. (A question or an instrument can measure something consistently wrong!) Validity *requires* Reliability. (If something can’t be measured accurately, with consistency, then there cannot be validity.)

V. Discussion of YOQL-R from assigned reading

**\*\*\*Student report: In the article “The cultural adaptation of the Youth Quality of Life Instrument” the authors use the term “cultural equivalence.” Define it, and describe a few of the components.**

**\*\*\*Student report: In the same article, researchers recommended changes not only to the Spanish version of the instrument, but also the English version. Why?**

A. Dimensions of instrument and examples of questions

VI. Small Group Exercise—Developing Questions and Establishing Reliability and Validity

**Research scenario:** You are working as a social worker in a criminal justice probation agency, and wonder what some of the factors are leading to repeated criminal involvement of those managed by your agency over the past five years who were convicted and then released. You can interview a sample of them. The literature review shows the importance of the following variables:

- Seriousness of first offense
- Educational attainment/completion
- Employment status after first offense
- Availability of family or social support
- Use of addictive substances

1. **Define two independent variables operationally.** Develop three closed-ended survey questions that address the operationalization of each variable. You may use any of these levels of measurement:

- a) Nominal variable (e.g. “Yes or “No” response)
- b) Ordinal variable (e.g. Likert scale categories)
- c) A continuous variable (indicate what higher and lower scores mean)

2. **Define the dependent variable operationally.** Develop three survey questions that address the operationalization of the variable. You may use any of these levels of measurement:

Dependent Variable:

- a) Nominal variable (e.g. “Yes or “No” response)
- b) Ordinal variable (e.g. Likert scale categories)
- c) A continuous variable (indicate what higher and lower scores mean)

**3. For each independent variable:**

- a) Identify one type of reliability that is relevant to this variable and its measurement. How would you ensure reliability?
- b) How would you establish either *criterion* or *construct* validity of this measure?

**Next Week**

Note: next week we will conclude material on question and survey instrument development and begin the Sampling material; we will also talk about the format for the literature review.

Study questions for next week:

1. What is the difference between the concepts *sample* and *population*? What do we mean by *generalizing to the population*?
2. What is the main difference between probability sampling and non-probability sampling procedures?
3. If you were to pick a type of sampling procedure for your project, which would it be?

Important Concepts

- Reliability
- Interrater reliability
- Test-retest reliability
- Alternate forms reliability
- Internal consistency reliability
- Coefficient alpha
- Validity
- Face validity
- Content validity
- Criterion-related validity (known-groups, predictive, concurrent)
- Construct validity (convergent, discriminant)