Tests and Scales/ Reliability and Validity

Measuring of concepts in public health research entails measuring something that is not tangible unlike measuring physical properties such as age, heath, or weight; therefore, a formalized definition must be created. The concept must be developed into an operational definition. According to Crosby, DiClemente, and Salazar (2006) once the concept is operationalized it becomes a construct which must be measured with multiple questions or items designed to be distinct “effect indicators” of the construct. For example, measuring self-esteem (a construct) multiple effect indicators can be used to measure it, the measure of self-esteem would be a scale. A scale is then a measure of correlated items (variables), while an index refers to a measure in which the items are heterogeneous and may not be correlated with each other, considered causal indicators influencing the value of the construct. The Apgar scale is an example of an index since it rates newborns on heterogeneous characteristics respiration, heart rate, reflex response, and muscle tone (Crosby, DiClemente, and Salazar, 2006).

Indexes and scales not only allow the researcher in public health to represent several variables by a single score, but provide the means to statistical analysis which increases the reliability of the measure itself (Frankfort-Nachmias & Nachmias (2008). Any scale or index measurement must show reliability and validity. Reliability: measure consistently provides the same answer every time is used (Frankfort-Nachmias & Nachmias (2008), it assesses whether the indicator consistently measures the concept (Novick, Morrow, and Mays, 2008). Validity: the measure is measuring what is supposed to measure (Crosby, DiClemente, and Salazar, 2006), is an indicator’s capacity to measure the intended concept (Novick, Morrow, and Mays, 2008).

Testing reliability of a measure can be performed by: test- retest method, computing the inter-item correlations between all items in the scale [(Pearson’s r bivariate correlation coefficient and/or Cronbach’s alpha statistic score (Frankfort-Nachmias & Nachmias (2008)), and by using the split-half method. Validity can be established by either face validity or content validity techniques. Face validity is judged by a panel of qualified professionals by assessing if the index or scale measures the construct. Scales and indexes can be assessed for content validity in a specific way for each. Scales content validity assessment is done based on the items representing the “universe” of all possible indicators relevant to the construct, while index content validity is assessed on viewing the items more as a consensus of items and are dependent on prior research and theory of the construct (Crosby, DiClemente, and Salazar, 2006).

In a study performed by Savoia et al (2009) titled “Assessing Public Health Capabilities during Emergency Preparedness Tabletop Exercises: Reliability and Validity of a Measurement Tool”, the authors developed a measurement tool; a 37-item questionnaire, to assess five public health function capabilities (leadership and management, mass casualty care, communication, disease control and prevention, and surveillance and epidemiology) to study population composed of 179 public health officials who attended three table top exercises in Massachusetts and Maine between September 2005 and November 2007. Savoia et al (2009) measurement tool was based
on a PHEP logic model used previously to examine the public health response to previous outbreaks in the U.S.; monkey pox, severe acute respiratory syndrome (SARS), West Nile virus, and hepatitis A. Three table top exercises were executed. After each tabletop scenario participants completed a questionnaire to access the community they served ability to respond focusing on a set of indicators.

The statistical package SPSS was used for statistical analysis to measure internal consistency of multi item-scales, as a measure of reliability, calculated by Cronbach’s alpha and correlations were determined by Pearson’s product moment coefficient. Validity was measured by comparing the tabletop participants’ responses to external evaluators’ responses using intra-class correlation coefficients (Savoia et al., 2009). The authors concluded that the reliability and validity of the instruments used were sample and context- specific and should be retested if the instruments are to be used in a different setting with a population with different characteristics of the one used for this study. Nevertheless, the tool used in this study is cost-effective, valuable, and provides a way to measure PHEP progress in similar communities.

References:


