



Measurements in Quantitative Research: How to Select and Report on Research Instruments

Teresa L. Hagan, BSN, BA, RN

Measures exist to numerically represent degrees of attributes. Quantitative research is based on measurement and is conducted in a systematic, controlled manner. These measures enable researchers to perform statistical tests, analyze differences between groups, and determine the effectiveness of treatments. If something is not measurable, it cannot be tested.

Some measures in nursing research can be directly quantified. For example, blood pressure can be measured with increasing precision using patient recall, blood pressure cuff, or an intra-arterial line. All of those measurements have a degree of error, but the concept of blood pressure can be measured with some degree of certainty. Other concepts in nursing research are dynamic and abstract, making direct measurement impossible. Rather, this type of research must depend on reports of actions, attitudes, or behaviors relevant to that concept. Social-psychological concepts require more creativity. Measuring subjective states or abstract concepts like depression, self-efficacy, and optimism must be measured by observing or asking participants about behaviors and attitudes that represent these concepts. Nursing research frequently uses self-report surveys to measure concepts critical to practice. Despite that, these concepts are difficult to operationalize (or make measurable).

Oncology nursing and research is not immune to such measurement problems. Two classic examples of such concepts are cancer-related fatigue (CRF) and quality of life (QOL). Capturing these concepts is necessary for oncology nursing research and practice; however, these concepts remain wrought with conceptual confusion and measurement imperfections. A comparison of fatigue instruments demonstrated low construct

validity among several instruments (Meek et al., 2000). McCabe and Cronin (2011) provided a thorough critique of health-related QOL instruments, arguing that frequently used instruments fail to include the most influential factors associated with the concept and lack clear meaning as outcome measures. For oncology nursing science to test the theoretical frameworks and conceptual models it intends to test (and ultimately improve patient and healthcare outcomes of those affected by cancer), the instruments used to quantify these concepts and others must be psychometrically appropriate and rigorous.

Developing and designing a research study requires significant time to define research questions, refine theoretical frameworks, and delineate study procedures. Choosing how to quantify the study's variables is, however, of utmost importance (Polit & Beck, 2012). This article aims to review issues regarding instrument selection and key components when reporting on study instruments used in a quantitative study.

The psychometric properties of instruments are primarily defined by the instrument's reliability and validity (Kimberlin & Winterstein, 2008). Reliability refers to the consistency of scores reported by a study participant. Validity refers to the accuracy of score interpretations. An important, yet often overlooked, distinction is made in these definitions. Rather than the instrument itself being reliable or valid, the scores' interpretations of that instrument are said to be reliable and valid. Although seemingly trivial, this distinction emphasizes the conditional nature of psychometric strength. Psychometric strength is not an unchanging quality of an instrument, but rather the population that is completing the instrument and its

respective scores earn the properties of reliable and valid (Soeken, 2005).

Reliability

Reliability can be measured multiple ways depending on the type of instrument (Polit & Beck, 2012). The most common forms include: (a) test-retest (comparing item responses from same participants at different time points), (b) internal consistency (comparing item responses against other item responses), and (c) scorer reliability (comparing one reviewer with another reviewer—in case a scorer is completing the instrument). If reliable, researchers can assume the instrument's scores are dependable, consistent, and more likely to be generalized to other samples, times, reviewers, and samples of behaviors. If inconsistent, then the error may be because of problems with the items or reviewers and will need to be examined and addressed. These problems must be addressed before evaluating the validity of score interpretation; validity cannot exist without reliability (Kimberlin & Winterstein, 2008).

Measures of reliability evaluate the extent of individual differences between scores across groups of respondents. One of the most commonly reported reliability measurements is the reliability coefficient. These statistics are based on correlations between scores either on the same test, equivalent tests, or along timepoints. The correlation calculates the variance of the true score divided by the observed score. The higher the correlation, the more the true and observed

ONF, 41(4), 431–433.
doi: 10.1188/14.ONF.431-433