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

- **What is the concept you are trying to operationalize?**

  Concepts usually are situated within conceptual models, and these models usually come from theoretical frameworks. Clarity in the theoretical and conceptual models being tested or explored can help clarify precisely what concepts should be measured. Then, researchers can look to see if this concept previously has been operationalized by a particular instrument.

- **What population do you wish to generalize to?**

  If your research is focused on children with diabetes, then instruments that were validated among adults with acute respiratory illness may not be appropriate.

- **What instruments do similar research studies use?**

  Equivalency among measurements facilitates syntheses of findings in the future.

- **What does the instrument look like?**

  Read the instructions, item questions, and response options. Consider the reading level and content of the questionnaire. Is it appropriate for your sample?

- **How acceptable is the instrument?**

  Will participants find the questions intrusive or irrelevant? This may cause a problem with missing or incomplete data.

- **How is the instrument administered?**

  Is the format self-administered, interview, or a mixture of both? Does this affect how many people can be included in the study? How simple and accurate will it be to interpret the score results?

- **How long does it take to complete? Is the number of questions excessive or burdensome?**

  Instruments with too many questions may cause participants to stress or become fatigued.

- **Are there copyrights or costs associated with the measure?**

  Contact the authors or people who own the instrument ahead of time to gain permission and discuss any fees associated with the administration and/or scoring of the instrument.

Figure 1. What to Consider When Selecting Instruments for a Quantitative Research Study

scores are similar; therefore, less error occurs. Long instruments (more than 40 items) will have an inflated coefficient because they have more items with which to be correlated.

Validity

Validity can be measured in multiple ways. If valid, researchers can be confident in the score interpretations and that the measurement is indeed measuring the desired concept. The process of establishing validity involves collecting various forms of evidence to support that the score interpretations are accurate. Before doing this, the researcher must clearly state what the score interpretations are: (a) What concept is being measured? (b) What are the intended interpretations of the instrument? And (c) What (if any) assumptions do the interpretations rely on?

Prior to the 1990s, psychometricians separated validity into content validity, criterion validity, and construct validity (Hubley & Zumbo, 1996). Many research articles still refer to this outdated nomenclature. Currently, rather than thinking of various types of validity as separate entities, each is thought to equally add evidence for the researcher's overall argument that the instrument captures what it claims to measure. This modern validity taxonomy includes various types of evidence based on the content of the measure compared to the construct of interest, internal structure (relationships between items compared to reported scores), and external structure (relationships between the survey results and external variables) (Messick, 1993).

Other types of evidence include content, internal structure, and external structure evidence. Content evidence refers to the relationship between the items in the instrument and the overall construct being measured. The instrument should be relevant and representative of the construct without extraneous items. Internal structure evidence evaluates the relationship between items relative to the underlying conceptual framework of the construct, and is usually performed with a factor analysis to uncover relationships among items, dimensions, and the overall scale. External structure evidence can be used by comparing scores with similar instruments and/or dissimilar instruments, or by measuring the degree to which an individual's behavior can be predicted based on scores from the instrument. Given the multiple types of validity evidence that can be argued, the researcher must select the most appropriate kinds of validity to test. DeVon et al. (2007) provided an informative, succinct description of multiple types of validity to assist researchers.

Selecting Instruments

The major concern when selecting an instrument is that it measures the concepts relevant to the research question. Does the instrument allow the researcher to measure the predictor, outcome, and/or mediator and moderator variables needed to answer the research question? Is it best suited to quantify the theoretical and conceptual framework? Other questions should be considered when selecting instruments for a research study (see Figure 1).

Finding reliable, valid instruments can be done by reading relevant research studies and finding what measurements other researchers in your field use. The U.S. National Library of Medicine offers a searchable database of research instruments (wwwcf.nlm.nih.gov/hsrr_search/index.cfm), as does the Agency for Healthcare Research and Quality (www.qualitymeasures.ahrq.gov).

Reporting Instruments

After a research study has been completed, results should include information about the selection, administration, and results of the instrument's performance. Even if the reliability and validity statistics and evidence were reported in the selection of the instruments, they should be calculated again to provide psychometric evidence for the present sample. If any measurement error, response bias, or coverage bias is found during the study, such sources of error should be reported because they can affect the instrument's score interpretations (DeVellis, 2012). In case the reliability statistics fall below set standards (e.g., Cronbach alpha < 0.8 or weaker validity as evidenced by unclear factor analysis), then this must also be reported because score interpretations will be directly affected.

Conclusion

Measurement, the cornerstone of scientific research, requires significant

forethought. Conceptually and statistically, the results of a study hinge on how the sample responds to the instruments being used. As nurses and scientists, this refresher on the importance of psychometric clarity, reliability, and validity can help researchers select instruments and report their instruments in ways that improve the quality and impact of data and knowledge.

Teresa L. Hagan, BSN, BA, RN, is a predoctoral fellow in the Department of Acute and Tertiary Care in the School of Nursing at the University of Pittsburgh in Pennsylvania. No financial relationships to disclose. Hagan can be reached at tlh42@pitt.edu, with copy to editor at ONFEditor@ons.org.

Key words: measurements; quantitative research; reliability; validity

References

DeVellis, R.F. (2012). *Scale development: Theory and applications* (3rd ed.). Washington, DC: Sage.

DeVon, H.A., Block, M.E., Moyle-Wright, P., Ernst, D.M., Hayden, S.J., Lazzara, D.J., . . . Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship, 39,* 155–164.

Hubley, A.M., & Zumbo, B.D. (1996). A dialectic on validity: Where we have been and where we are going. *Journal of General Psychology, 123,* 207–215.

Kimberlin, C.L., & Winterstein, A.G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacists, 65*(1), 2276–2284.

McCabe, C., & Cronin, P. (2011). Issues for researchers to consider when using health-related quality of life outcomes in cancer research. *European Journal of Cancer Care, 20,* 563–569.

Meek, P.M., Nail, L.M., Barsevick, A., Schwartz, A.L., Stephen, S., Whitmer, K., . . . Walker, B.L. (2000). Psychometric testing of fatigue instruments for use with cancer patients. *Nursing Research, 49,* 181–190.

Messick, S. (1993). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Phoenix, AZ: Oryx Press.

Polit, D.F., & Beck, C.T. (2012). *Nursing research: Generating and assessing evidence for nursing practice* (9th ed.). Philadelphia, PA: Lippincott Williams and Wilkins.

Soeken, K.L. (2005). Validity of measures. In C.F. Waltz, O.L. Strickland, and E.R. Lenz (Eds.), *Measurement in nursing health research* (3rd ed., pp. 154–189). New York, NY: Springer.