



MEASUREMENT in

NURSING AND  
HEALTH  
RESEARCH

Carolyn Feher Waltz  
Ora Lea Strickland  
Elizabeth R. Lenz

SPRINGER  PUBLISHING COMPANY

FOURTH EDITION

**Measurement in  
Nursing and  
Health Research**

*Fourth Edition*

**Carolyn Feher Waltz, PhD, RN, FAAN**, is a Professor at the University of Maryland School of Nursing in Baltimore, Maryland. She received her BSN and MS degrees in public health nursing from the University of Maryland and her PhD in education with specialization in research, evaluation, and measurement from the University of Delaware. Dr. Waltz is an internationally recognized expert in the development of outcomes evaluation and measurement in clinical and educational settings and an adjunct member of GRIISIQ, Quebec Interuniversity Nursing Intervention Research Group, a consortium of researchers from all Canadian universities in Quebec. She has published extensively including award-winning books and articles on research, measurement, nursing outcomes, program development, and evaluation. She has received research funding from multiple sources, authored numerous articles, books and book chapters, and was a founding editor of the journal *Evaluation & the Health Professions*. She has provided numerous clinical and educational outcomes evaluations, consultations, presentations, and workshops to varied health care audiences nationally and internationally and is widely acknowledged as a pioneer in outcomes measurement, evaluation, and research.

**Ora Lea Strickland, PhD, RN, FAAN**, is Professor in the Nell Hodgson Woodruff School of Nursing at Emory University in Atlanta, Georgia, with specialty in child development and family relations. She is the founder and senior editor of the *Journal of Nursing Measurement*. As an internationally known specialist in nursing research, measurement, evaluation, maternal and child health, and parenting, Dr. Strickland has published widely in professional journals and books, and is called upon as a consultant by health care agencies, universities, government agencies, and community organizations. She has presented more than 300 lectures, speeches, and workshops, and her research has been featured in newspapers and magazines, as well as on radio and television.

**Elizabeth R. Lenz, PhD, RN, FAAN**, is Dean and Professor at the Ohio State University College of Nursing. Prior to assuming the deanship in September 2001, she held faculty and academic administrative positions at Columbia University, the Pennsylvania State University, the University of Maryland, Georgetown University, and Boston College. She is known for her work in doctoral nursing education and her nursing research career has focused on the family as the context for health and illness. Recent research concerns the outcomes and quality of nurse practitioner practice. She is a former member of the editorial board of *Nursing Outlook*, and between 1999 and 2001 she served as an editor of *Scholarly Inquiry for Nursing Practice*.

# **Measurement in Nursing and Health Research**

*Fourth Edition*

**Carolyn Feher Waltz, PhD, RN, FAAN**  
**Ora Lea Strickland, PhD, RN, FAAN**  
**Elizabeth R. Lenz, PhD, RN, FAAN**

  
SPRINGER PUBLISHING COMPANY  
New York

Copyright © 2010 Springer Publishing Company, LLC

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Springer Publishing Company, LLC, or authorization through payment of the appropriate fees to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400, fax 978-646-8600, info@copyright.com or on the web at www.copyright.com.

Springer Publishing Company, LLC  
11 West 42nd Street  
New York, NY 10036  
www.springerpub.com

*Acquisitions Editor: Allan Graubard*  
*Project Manager: Laura Stewart*  
*Cover Design: Steve Pisano*  
*Composition: Apex CoVantage, LLC*

E-book ISBN: 978-0-8261-0508-0

09 10 11 12 / 5 4 3 2 1

The author and the publisher of this Work have made every effort to use sources believed to be reliable to provide information that is accurate and compatible with the standards generally accepted at the time of publication. Because medical science is continually advancing, our knowledge base continues to expand. Therefore, as new information becomes available, changes in procedures become necessary. We recommend that the reader always consult current research and specific institutional policies before performing any clinical procedure. The author and publisher shall not be liable for any special, consequential, or exemplary damages resulting, in whole or in part, from the readers' use of, or reliance on, the information contained in this book. The publisher has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

---

### Library of Congress Cataloging-in-Publication Data

Waltz, Carolyn Feher.

Measurement in nursing and health research / Carolyn Feher Waltz, Ora Lea Strickland,  
Elizabeth R. Lenz. — 4th ed. p. ; cm.

Includes bibliographical references and index.

ISBN 978-0-8261-0507-3 (alk. paper)

1. Nursing—Research—Methodology. 2. Medical care—Research—Methodology. 3. Measurement.  
I. Strickland, Ora. II. Lenz, Elizabeth R., 1943– III. Title.

[DNLM: 1. Nursing Research—Nurses' Instruction. 2. Health Services Research—Nurses' Instruction.  
3. Research Design—Nurses' Instruction. WY 20.5 W241m 2010]

RT81.5.W36 2010

610.73072—dc22 2010003793

---

Printed in the United States of America by Bang Printing.

---

# Contents

<i>Contributors</i>	<i>vii</i>
<i>Preface</i>	<i>ix</i>
<i>Acknowledgment</i>	<i>xi</i>

## **Part I Basic Principles of Measurement**

1 Introduction to Measurement	3
2 Operationalizing Nursing Concepts	27
3 Measurement Theories and Frameworks	49

## **Part II Understanding Measurement Design**

4 Strategies for Designing Measurement Tools and Procedures	91
5 Measurement Reliability	145
6 Validity of Measures <i>With Karen L. Soeken</i>	163
7 Standardized Approaches to Measurement	203
8 Measurement Issues in Qualitative Research <i>Judith E. Hupcey</i>	225

## **Part III Measurement via the Digital World**

9 Using Existing Administrative and National Databases <i>Meg Johantgen</i>	241
10 Computer-Based Testing <i>Louise S. Jenkins</i>	251
11 Internet Data Collection	259

**Part IV Instrumentation and Data Collection Methods**

12	Observational Methods	271
13	Content Analysis	279
14	Interviews	287
15	Questionnaires	301
16	The Delphi Technique	311
17	Visual Analog Scales	319
18	Magnitude Estimation Scaling	325
19	Guidelines for Writing Multiple-Choice Items <i>Karen L. Soeken</i>	329
20	Measurement of Physiological Variables Using Biomedical Instrumentation <i>Kathleen S. Stone and Susan K. Frazier</i>	335
21	Evaluation of Measurement Precision, Accuracy, and Error in Biophysical Data for Clinical Research and Practice <i>Nancy A. Ryan-Wenger</i>	371
22	Collecting Sensitive Information	385
23	Selection and Use of Existing Instruments	393

**Part V Measurement Issues**

24	Ethical Issues	401
25	Issues in Measuring Quality Care <i>Karen Beckman Pace</i>	417
26	Other Measurement Issues	433
	<i>Index</i>	477

---

# Contributors

**Karen Beckman Pace, PhD, RN**

Senior Program Director  
National Quality Forum  
Washington, DC

**Susan K. Frazier, PhD, RN**

Associate Professor  
University of Kentucky  
College of Nursing  
Lexington, Kentucky

**Judith E. Hupcey, EdD, RN, CRNP**

Associate Professor  
Pennsylvania State University  
School of Nursing  
Hershey Medical Center  
Hershey, Pennsylvania

**Louise S. Jenkins, PhD, RN**

Professor  
University of Maryland  
School of Nursing  
Baltimore, Maryland

**Meg Johantgen, PhD, RN**

Associate Professor  
University of Maryland  
School of Nursing  
Baltimore, Maryland

**Nancy A. Ryan-Wenger, PhD, RN, CPNP, FAAN**

Director of Nursing Research  
Nationwide Children's Hospital  
Investigator—Center for Innovation in  
Pediatric Practice  
Research Institute at Nationwide Children's  
Hospital  
Columbus, Ohio

**Karen L. Soeken, PhD**

Professor Emeritus  
University of Maryland  
School of Nursing  
Baltimore, Maryland

**Kathleen S. Stone, PhD, RN, FAAN**

Professor Emeritus  
The Ohio State University  
College of Nursing  
Columbus, Ohio

*This page intentionally left blank*

---

# Preface

The fourth edition of this text was written taking into account the needs and interests of nurses and our expanded audience of health researchers, the significant increase in interdisciplinary research collaboration, and growing emphasis on evidence-based practice as an important research outcome across all health disciplines. As in earlier editions, the intent, within the context of these trends, is to present a pragmatic account of the process involved in designing, testing, selecting and/or evaluating instruments, methods, and procedures for measuring variables in clinical, educational, and research settings.

We continue to strive to meet the needs of a large and diversified audience ranging from neophytes to those with more advanced knowledge and experience in measurement. Thus, we do not assume that most readers have an extensive background and experience in measurement or statistics. Rather, we begin our discussion of content assuming little background in these areas and subsequently develop, explain in detail, and illustrate by example the concepts and principles operationally important to the content presented. In this manner, it is possible for the less sophisticated reader to develop the level of knowledge necessary to understand the content that is included for the benefit of the more advanced reader. Thus, this book should serve as a valuable resource for readers who seek basic and advanced content to develop their skill in measurement.

This edition includes the best of the measurement content, strategies, and procedures presented in the previous editions. In addition, it provides the most up-to-date content, strategies, and procedures available with direct applicability for nurses and health researchers in a variety of roles including those of student, educator, clinician, researcher, administrator, and

consultant. As in previous editions, the focus is on increasing the reader's ability to employ measures that are operationalized within the context of theories and conceptual frameworks, derived from sound measurement principles and practices, and adequately tested for reliability and validity using appropriate methods and procedures. Throughout this edition, examples and studies conducted by nurses and health researchers in varied settings and/or in collaboration with others in their own and other disciplines are provided to illustrate important content and to reinforce the importance of using sound measurement principles and practices within the context of evidence-based practice. Attention is given to measurement issues resulting from changes in nursing and health research that have increased in number and complexity since the time of the last edition. Additional reference sources, readily available in libraries and/or online, are provided for readers who desire to further pursue the topics presented. Whenever possible, comprehensive summaries of literature in an area of interest are cited rather than a myriad of individual books and articles.

Some key features of this edition include:

- Chapter 6, "Validity of Measures," provides an updated and expanded reconceptualization and definition of measurement validity, and includes new strategies for testing validity of measures.
- The content of Chapter 8, "Measurement Issues in Qualitative Research," has been expanded to include additional content relevant to the use of a mixed methods approach to enhance measurement outcomes.
- A new section, "Part III: Measurement via the Digital World," has been added

and includes Chapter 9, “Using Existing Administrative and National Databases,” which provides examples of major databases and focuses on (1) using large data sets from national longitudinal research as sources for nursing and health research, and (2) employing large data sets from clinical information systems in hospital and multifacility health systems, especially in addressing quality and safety; Chapter 10, “Computer-Based Testing,” which focuses on the use of the Internet in the conduct of research and measurement efforts, including identifying issues and how to address them; and Chapter 11, “Internet Data Collection,” which contains expanded content on issues of data privacy and how to address them.

- Chapter 20, “Measurement of Physiological Variables Using Biomedical Instrumentation,” has been expanded and is focused on the methods for measuring physiological variables using biomedical instrumentation, including issues to be addressed and use in translational research.
- There is a new Chapter 21, “Evaluation of Measurement Precision, Accuracy, and Error in Biophysical Data for Clinical Research and Practice,” which presents content regarding the application of measurement principles and practices in clinical research and practice that focuses

on clinical data collection methods, including (1) clinimetrics and the unique challenges of conducting research in clinical settings and situations where strict experimental control and standardization are not optimal and how to address them, (2) standards for evaluating precision and accuracy of clinical measures, and (3) issues to be addressed relative to use of outcomes as a basis for clinical practice.

- There is a new Chapter 25, “Issues in Measuring Quality Care,” which focuses on principles and practices in quality measurement, identifies issues in measuring quality of care, and explores challenges and how to address them.

We appreciate the positive feedback we have received from satisfied readers who have valued previous editions as a textbook for research and measurement courses, and from those conducting their own research and/or seeking to evaluate the quality of others’ research to decide whether to use their results as a basis for practice. Also, we thank the many readers who requested another edition and sincerely hope that you will find this fourth edition a significant resource.

CAROLYN FEHER WALTZ, PhD, RN, FAAN  
ORA LEA STRICKLAND, PhD, RN, FAAN  
ELIZABETH R. LENZ, PhD, RN, FAAN

---

# Acknowledgment

The authors sincerely appreciate the contributions of Sandra A. Murphy whose commitment, perseverance, organizational ability, and many talents were instrumental to us in preparing this manuscript. Sandy, we can't thank you enough for so graciously contributing your time and effort to ensure the completion of this book. You are the best!

*This page intentionally left blank*

# **Part I**

---

## **Basic Principles of Measurement**

*This page intentionally left blank*

# Introduction to Measurement

The use of sound measurement principles and practices is an essential component of well-designed research studies and is of utmost importance when the goal is to employ research results as a basis for practice. In this chapter, we discuss terms and ideas essential to understanding the content of subsequent chapters, present an overview of the types of measures most often used, and introduce the two essential characteristics of any measuring tool or method, reliability and validity.

## THEORY-BASED MEASURES

A *theory* is defined as “a set of interrelated constructs, definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting the phenomena” (Kerlinger, 1973, p. 9). Polit and Beck (2008) note that a theory “is a systematic, abstract explanation of some aspect of reality. In a theory, concepts are knitted together into a coherent system to describe or explain some aspect of the world” (p. 57). A *theoretical rationale* according to LoBiondo-Wood and Haber (1994) “provides a road map or context for examining problems, and developing and testing hypotheses. It gives meaning to the problem and study findings by summarizing existing knowledge in the field of inquiry and identifying linkages among concepts” (p. 157). *Conceptual models*, they note, provide a context for constructing theories that deal with phenomena of concern to a discipline and help define how it is different from other disciplines. Polit and Beck (2008) use the terms *conceptual model* and *conceptual framework* interchangeably. In their view a conceptual model includes “interrelated concepts or abstractions assembled

together in a rational scheme by virtue of their relevance to a common theme” (p. 749).

For example, Arwood and Kaakinen (2009) designed a simulation model based on language and learning that was evolved from Arwood’s Neuro Sematic Language Learning Theory that provides a hierarchical framework for assessing and measuring conceptual learning outcomes. Evans and colleagues (2009) employed the Biopsychosocial Model to understand the health benefits of Yoga. They present a review and conceptual model of the potential biopsychosocial benefits of Yoga and empirical evidence that supports the involvement of physical, psychological, and spiritual domains as possible mechanisms of actions of Yoga upon well-being. Krogh and Naden (2008) report on a project undertaken by them using an information management system to develop a conceptual framework that would provide clinicians with an approach to documentation consistent with legal and organizational requirements and enable them to retain the ability to record all aspects of clinical nursing. The result was the Norwegian documentation KPO model that incorporates the Nursing Minimum Data Set and NANDA (Nursing Diagnosis, Nursing Intervention Classification, and Nursing Outcome Classification) and focuses on quality assurance, problem solving, and caring.

Various authors have defined and analyzed concepts in a number of ways (Allegood & Marriner-Tomey, 2002; George, 2002; Glaser & Strauss, 1967; Leonard & George, 1995; Meleis, 1999; Orem, 1995; Polit & Beck, 2008; Renpenning & Taylor, 2003). Simply defined, a *concept* is a thought, notion, or idea. It is an abstraction. For example, nursing concepts are thoughts, notions, or ideas about nursing or nursing practice. Thus, concepts define the content of interest in measuring phenomena. *Phenomena*

are observable facts or events. To render concepts measurable, it is necessary to translate them into phenomena. When one operationalizes a concept, one translates an abstract concept into concrete observable events or phenomena. For example, the concept of “attitude” is frequently operationalized as a tendency to respond in a consistent manner to a certain category of stimuli (Campbell, 1963). If the stimulus is a 17-item questionnaire regarding children’s concerns about health care, such as the Child Medical Fear Scale developed by Broome and Mobley (2003, pp. 196–206), in which the child indicates whether he or she was “not at all,” “a little,” or “a lot” afraid of selected experiences associated with health care described in the items, and the subject responds “a lot” to the majority of the 17 items, one would infer from the responses that the child’s attitude or fearfulness of the experiences associated with health care was high.

*Variables* are quantities that may assume different values; they are changeable. The process whereby one decides how to measure a variable is referred to as *instrumentation*, that is, the process of selecting or developing tools and methods appropriate for measuring an attribute or characteristic of interest. In the example above, the 17-item questionnaire was a form of instrumentation selected to measure attitudes of children toward the medical experience. Instrumentation is a component of the measurement process. *Measurement* is defined as the process of assigning numbers to objects to represent the kind and/or amount of attributes or characteristics possessed by those objects. This definition of measurement includes what has traditionally been referred to as *qualitative measurement* (i.e., assigning objects to categories that represent the kind of characteristic possessed and that are mutually exclusive and exhaustive) as well as *quantitative measurement* (i.e., assigning objects to categories that represent the amount of a characteristic possessed).

The utilization of a conceptual framework to systematically guide the measurement process increases the likelihood that concepts and variables universally salient to nursing and health care practice will be identified and explicated. That is, when measurement concerns emanate from an empirical rather than a conceptual point of

view, there is higher probability of investigating and measuring these variables from an esoteric or limited perspective that overlooks important dimensions of the variables that should be measured. Concepts of interest to nurses and other health professionals are usually difficult to operationalize, that is, to render measurable. This is explained in part by the fact that nurses and other health professionals deal with a multiplicity of complex variables in diverse settings, employing a myriad of roles as they collaborate with a variety of others to attain their own and others’ goals. Hence, the dilemma that they are apt to encounter in measuring concepts is twofold: first, the significant variables to be measured must somehow be isolated; and second, very ambiguous and abstract notions must be reduced to a set of concrete behavioral indicators. What tools, therefore, are available to nurses and other health care professionals who must begin to grapple with this dilemma?

Because of the increased interest in evidence-based practice across health care disciplines and the challenge to provide services of broadening scope and diversity in order to keep pace with the rapidly changing and volatile health care scene coupled with severe shortages in nursing and other health care disciplines, a great deal of controversy and confusion has ensued regarding which functions should be included within the realm of each of the practice disciplines and which information should be shared. For nursing, this is evidenced by the proliferation of definitions and models for nursing practice evident in the literature. Although, for the most part, the definitions of nursing advanced in the literature remain ambiguous and global, in each view the major focus for nursing practice can be placed on a continuum ranging from direct to indirect involvement in patient care. Direct nursing practice involves the continuous, ongoing provision of direct services to patients and clients (e.g., the primary care nurse practitioner provides direct nursing services). Indirect nursing practice is usually characterized by activities on behalf of the patient, that is, working with or through others who are directly responsible for the provision of direct services to patients and clients. Nursing education, administration, and health policy activities exemplify indirect nursing practice.

This scheme for categorizing nursing practice has utility for the nurse who is attempting to operationalize nursing concepts.

More specifically, the first task is to identify a definition of nursing that is consistent with the nurses' own views and beliefs about nursing practice. Similarly, although the extent to which available conceptual frameworks and models for nursing practice have been refined and tested varies, their very existence affords nurses a rich opportunity to select a conceptual framework to guide them in systematically identifying and further explicating concepts and variables germane to nursing and nursing practice concerns within their primary focus. The problems, settings, roles, and purposeful activities undertaken by nurses will differ, depending upon whether their primary focus is direct or indirect nursing practice. Furthermore, the goals for, and outcomes likely to result from, the application of direct and indirect nursing processes will vary. Although there will be differences among each of these categories of practice, there will also be commonalities in process and outcomes within each focus. Therefore, if nurses consider their measurement concerns within the context of their primary focus, delimit the processes and outcomes that characterize that practice, and then search for behavioral indicators within the primary focus that extend beyond their immediate setting (i.e., that are common across settings similar to their own), they are apt to reduce the abstract concepts emanating from their conceptual framework to behavioral indicators with more universal acceptance than those likely to result from a more esoteric approach. In this manner, nurses will ultimately make a contribution to the profession by accruing information to add to the body of knowledge about nursing, the specific definition and conceptual framework employed, and its utility as a guide for operationalizing nursing and nursing practice concerns. It should be noted that when nurses whose measurement concerns emanate from their ongoing practice fail to step back and rethink the problem from a conceptual point of view, they also have a high probability of investigating and measuring their variables from a limited perspective that overlooks important dimensions of the variables that should be measured.

For example, Stacey and colleagues (2009) report on their efforts to integrate a patient decision support theoretical framework and associated evidence-based resources throughout a baccalaureate nursing curriculum. Rew, Grady, Whittaker, and Bowman (2008) employed Social Cognitive Theory as a basis for their study to determine the effects of duration of homelessness and gender on personal and social resources, cognitive-perceptual factors, and sexual health behaviors among homeless use. The work of Ellenbecker, Porell, Samia, Byleckie, and Milburn (2008) exemplifies testing of a theoretical model of home health care nurse retention. The use of conceptual models as a basis for measurement is illustrated by the work of Landis, Parker, and Dunbar (2009) who employed a biobehavioral-ecological framework in their study of sleep, hunger, satiety, food cravings, and caloric intake in adolescents and Avci and Kurt (2008) who employed the Health Belief Model in their study of health beliefs and mammography rates of Turkish women living in rural areas. A conceptual approach within a qualitative phenomenological study was undertaken by Melby, Dodgson, and Tarrant (2008) who sought to describe the experiences of Western expatriate nursing educators teaching in East Asia. Additional examples of theories, conceptual frameworks, and theory-based measures can be found in Bramadat and Driedger (1993); Kempen, Miedema, Ormel, and Molenaar (1996); Fry (2001); Armitage (2001); Weber, Kopelman, and Messick (2004); March and Olsen (2006); Kaji, Koenig, and Bey (2006); Olsen (2007).

In nursing and the health professions a salient concern is often with the measurement of *process variables* such as with developing and implementing process measures of quality, which are dynamic, as well as *outcome variables*, which are usually static, and in which results of the measurement are likely to be applied to the solution of significant problems across practice settings as in evidence-based practice (EBP). For example, Rubin, Pronovost, and Diette (2001) recognize the importance when developing process measures of rigorously developing quality indicators "that will provide insights into opportunities to improve quality of care" and identify the following steps to be

taken. “Developing a process measure includes defining the purpose of and audiences for the measures, choosing the clinical area to evaluate, organizing the assessment team, choosing the component of the process to measure, writing the indicator specifications, performing preliminary tests of feasibility, reliability and validity, and determining scoring and analytical specifications” (pp. 489–496). The measurement of outcome variables is exemplified in the work of Schneider, Barkauskas, and Keenan (2008) who conducted a study to determine the sensitivity and responsiveness of the Outcome and Assessment Information Set (OASIS) and the Nursing Outcomes Classification (NOC) to the effects of home health care nursing intervention (p. 76).

Two salient characteristics of all measurement efforts are reliability and validity. First and foremost, tools and methods selected or developed for measuring a variable of interest must demonstrate evidence for reliability and validity. *Reliability* in this case refers to the consistency with which a tool or method assigns scores to subjects. *Validity* refers to the determination of whether a tool or method is useful for the purpose for which it is intended, that is, measures what it purports to measure. Second, in addition to the concern with instrument reliability and validity, attention needs to be given to the reliability and validity of the measurement process per se. To increase the probability that the measurement process will yield reliable and valid information, it is necessary whenever possible to employ multiple tools or methods to measure any given variable (all of which have demonstrated evidence for instrument reliability and validity) and to obtain information about any given variable from a number of different perspectives or sources.

For example, Aitken et al. (2008) conducted a study “to describe strategies used to administer an international multicenter trial to assess the effectiveness of a nursing educational intervention” (p. 101). Noting the importance of clinical relevance and that “in order to support appropriate development of an evidence base for practice, nursing interventions should be tested in multiple settings,” they employed multiple sites, countries, and methods (p. 101). Lee, Chaboyer, and Wallis (2008), in their study

to predict health-related quality of life three months after traumatic injury, employed multiple measurement methods including telephone interviews, clinical data, and questionnaires using rating scales.

Measurement reliability and validity is thus largely a function of a well-designed and well-executed measurement process. For this reason, the intent of this book is to provide the reader with a sound background in the theories, principles, and practices of measurement and instrumentation that are germane to the measurement of concepts in nursing and the health professions.

## MEASUREMENT FRAMEWORKS

Just as it is important to identify and employ a conceptual framework for determining what is to be measured and delineating how it will be operationalized, it is equally important to identify and employ a measurement framework to guide the design and interpretation of the measurement per se. The two major frameworks for measurement are the norm-referenced and the criterion-referenced approaches.

A *norm-referenced approach* is employed when the interest is in evaluating the performance of a subject relative to the performance of other subjects in some well-defined comparison or norm group. The Stress of Discharge Assessment Tool (SDAT-2) developed by Toth (2003, pp. 99–109) is an example of a 60-item norm-referenced measure of the stress experienced by patients at hospital discharge and in the early recovery period at home following acute myocardial infarction. Scores on each item in the SDAT-2 range from 1 to 5 points, depending on the patient’s degree of agreement with the item. A high score indicates high stress for that item, and a low score indicates low stress. The total possible score ranges from 60 to 300 points, and its value for a given subject takes on meaning when it is considered in light of the scores obtained by other patients who responded to the same tool.

Similarly, the results of the application of physiologic measures such as blood pressure readings are often interpreted on the basis of

readings (usually ranges of values) considered normal for some well-defined comparison group (e.g., Black males over 40 years of age with no significant health problems). It should be noted that the terms “norm-referenced” and “standardized” are not synonymous. *Standardized tests* are one type of norm-referenced measure; there are other types as well. Unlike most other norm-referenced measures, a standardized measure is designed by experts for wide use and has prescribed content, well-defined procedures for administration and scoring, and established norms. The Graduate Record Examination (Stein & Green, 1970), the National League for Nursing Achievement Test Battery (Waltz, 1988), and nurse practitioner certification examinations such as the Neonatal Intensive Care Nursing Examination (Perez-Woods, Burns, & Chase, 1989) are examples of standardized measures.

A key feature of a *norm-referenced measure* is variance. The task when using a norm-referenced measure is to construct a tool or method that measures a specific characteristic in such a way that it maximally discriminates among subjects possessing differing amounts of that characteristic, that is, spreads people out along the possible ranges of scores. For example, if the characteristic to be measured is knowledge of human sexuality content, then test items are designed to differentiate between individuals with varying levels of knowledge of the content. The goal is to obtain scores in such a manner that the result is a few high scores, most scores in the middle range, and a few low scores. If this goal is achieved, the resulting distribution of scores on the measure should look much like a normal curve. Figure 1.1(a) illustrates the distribution of scores that one would expect to result from the employment of the hypothetical 5-item, norm-referenced measure of human sexuality content.

The sole purpose of a *criterion-referenced measure* is to determine whether a subject has acquired a predetermined set of target behaviors. The task in this case is to specify the important target behaviors precisely and to construct a test or measure that discriminates between those subjects who have and those who have not acquired the target behaviors. How well a subject's performance compares with the performance of others is irrelevant when a criterion-referenced

framework is employed. Criterion-referenced measures are particularly useful in the clinical area when the concern is with the measurement of process and outcome variables.

For example, a criterion-referenced measure of process would require that one identify standards for the patient care intervention and then compare subjects' clinical performance with the standards for performance (i.e., predetermined target behaviors) rather than compare subjects' performance with that of other subjects, all of whom may not meet the standards. Standards might be those resulting from international and/or national professional organizations such as Guidelines for the Management of Patients with Chronic Stable Angina published by the American College of Cardiology and the American Heart Association in 2007 or local site-specific standards such as those agreed upon by health care providers serving on the evidence-based practice committee in a local oncology outpatient setting regarding the management of chemotherapy-induced nausea and vomiting in their setting. Similarly, when a criterion-referenced measure is employed to measure patient outcomes, a given patient's status is determined by comparing his or her performance with a set of predetermined criteria (e.g., EKG normal, diastolic pressure below 80, other vital signs stable for 4 hours post-op) or target behaviors (e.g., requests for pain medication have ceased by 2 days post-op, desire to return to normal activities is verbalized by 3rd day post-op) rather than by comparing his or her performance with that of other patients.

King's (2003, pp. 3–20) Measure of Goal Attainment Tool is an example of a criterion-referenced measure of functional abilities and goal attainment behaviors. This tool was constructed to assess individuals' physical abilities to perform activities of daily living and the behavioral response of individuals to the performance of these activities. Goal attainment was defined as mutual goal setting by nurse and patient, and assessment of goal attainment. Each of these areas comprises three components: (1) personal hygiene, (2) movement, and (3) human interaction. Essential tasks in performing actions related to each of these areas are reflected in items contained on the tool. Percentage scores are determined to evaluate the independence

or dependence of subjects related to the essential tasks. Thus, the individual's performance, as reflected by the score, is interpreted on the basis of his or her ability to perform the essential tasks and is not related to the evaluation of the performance of others using the same tool.

One would expect the distribution of scores resulting from a 5-item, criterion-referenced measure to look like the one illustrated in Figure 1.1(b). It should be noted that not only does the distribution of scores resulting from a criterion-referenced measure have less variance or spread than that resulting from a norm-referenced measure, but it also is skewed in shape. In a skewed distribution, scores tend to cluster at one end of the scale; in the example in Figure 1.1(b), the high end. A more detailed discussion of score spread

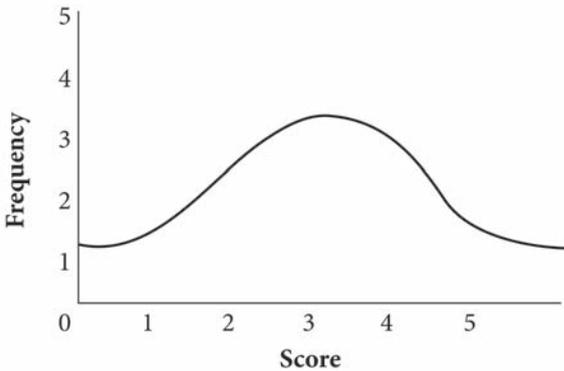
(variance) and distribution shape (normal and skewed) is presented in Chapter 3.

Because the design, scoring, interpretation, reliability, and validity testing for norm-referenced and criterion-referenced measures differ, it is important to decide which of the two measurement frameworks will be employed prior to the conduct of any other steps in the measurement process.

## TYPES OF MEASURES

Once the conceptual basis for the measure and the measurement framework has been determined, attention turns to the selection of the specific type of measure to be employed.

### A. Normal distribution of scores on a 5-item norm-referenced measure



### B. Positively skewed distribution of scores on a 5-item criterion-referenced measure

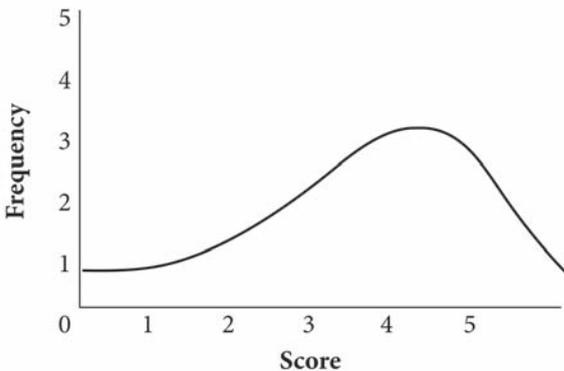


FIGURE 1.1 Distribution scores on (a) norm-referenced and (b) criterion-referenced measures.

Semantics is often a problem for newcomers to the field of measurement. Many different terms are employed to label like measures. For this reason, it is important to consider some of the classification schemes and terms that are used to label the different types of measures. In addition to being categorized as either norm-referenced or criterion-referenced, approaches to measurement are also usually referred to as qualitative or quantitative in nature. Whether the approach is qualitative, quantitative, or mixed methods will largely be a function of the measurement problem, the conceptual basis guiding the determination of what is to be measured, and the nature of the variables to be measured.

When a *qualitative approach* to measurement is employed, objects are assigned to mutually exclusive and exhaustive categories that represent the kind of characteristic possessed by the object. Qualitative measurement methods are usually indicated when the theoretical orientation for the measurement derives from phenomenology, existentialism, symbolic interactionism, or ethnomethodology (Chenitz & Swanson, 1986, p. 3). Knafl (1989) suggests that within nursing, ethnography, phenomenology, and grounded theory appear to be the dominant approaches. The major goal of qualitative methods is to document and interpret as fully as possible the whole of what is being measured from the frame of reference of the subjects involved (Diekelmann, 1992; Duffy, 1987; Filstead, 1974; Girot, 1993). Hinds and Young (1987) note that qualitative methods usually attempt to describe processes and thus tend to be measures of dynamic and changing phenomena. For example, Turrís (2009) employed a Grounded Theory approach in a study of women's decisions to seek treatment for the symptoms of potential cardiac illness. Data for the study were obtained from in-depth interviews with women who went to one of two emergency departments within a specific time period for the treatment of symptoms indicating potential cardiac arrest. Hendrickson (2008) conducted a study to explore worries, safety behaviors, and perceived difficulties in keeping children safe at home in a sample of low-income predominately non-English-speaking mothers as a foundation for later nursing intervention. The study employed a qualitative descriptive

design with content analysis of responses to three semi-structured interview questions (p. 137). A phenomenological approach was employed by Melby et al. (2008) to describe the lived experience of English-speaking Western nurse educators teaching in East Asian countries. Subjects were interviewed about their experiences and resulting narrative data were analyzed by each member of the research team "to identify themes and then through group discussions to reach consensus, develop clear understandings of language nuances and maintain the participant's voices" (p. 176). Mbweza, Norr, and McElmurry (2008) used a Grounded Theory approach to examine "the decision-making processes of husband and wife dyads in matrilinear and patrilinear marriage traditions of Malawi in the areas of money, food, pregnancy, contraception and sexual relations" (p. 12).

While the measurement strategies and techniques employed across the qualitative theoretical perspectives may differ somewhat, the types of measures employed are often the same and are also employed when quantitative approaches are used. Qualitative measurement methodologies generally include content analysis of documents; reviews of the literature and of findings from studies to identify common themes; participant and nonparticipant observations; interviews or focus groups that may be structured or unstructured, but are usually open-ended; and open-ended questionnaires. Qualitative data collection and analysis procedures are discussed in Chapter 8.

*Quantitative measurement* assigns objects to categories that represent the amount of a characteristic possessed by the object. Quantitative methods emphasize the search for facts and causes of human behavior through objective, observable, and quantifiable data (Duffy, 1987). Hinds and Young (1987) suggest that quantitative approaches provide outcome data and information on the representativeness of the studied sample and thus tend to be measures of more stable phenomena. Specific types of methods employed with quantitative approaches include the variety of types of measures discussed in later sections of this chapter.

Single studies of a problem using qualitative or quantitative methods rarely involve full

exploration of the problem area (Bergstrom, 1989). When a *mixed method approach*, that is, when a combination of both qualitative and quantitative methods are employed together, the two approaches to measurement provide information regarding the internal and external validity of the studies or measurement processes (Campbell & Fiske, 1959; Vedich & Shapiro, 1955; Webb, Campbell, Schwartz, & Sechrest, 1966). For example, Escoffery, Glanz, Hall, and Elliott (2009) employed a multimethod process evaluation for a skin cancer prevention diffusion trial. Data collection methods in their study included surveys, database tracking, activity logs, process evaluation interviews, and site visits. Thus, to develop an adequate and useful repertoire of measurement principles and practices, one needs to understand when and how to use both qualitative and quantitative approaches to measurement, as well as how they can be combined as measurement methodologies.

Stainbeck and Stainbeck (1984) note that qualitative and quantitative approaches, because they derive from different perspectives, have several inherent differences that should be understood prior to considering how the two methods can complement each other in the measurement of nursing phenomena. Major differences noted by them include the following:

1. When quantitative approaches are employed, the goal is to arrive at an understanding of a phenomenon from the outsider's perspective by maintaining a detached, objective view that hypothetically is unbiased. The perspective, on the other hand, when qualitative approaches are employed is that of an insider and the goal is to obtain information by talking to and/or observing subjects who have experienced firsthand the phenomena under scrutiny.
2. Quantitative methods focus on the accumulation of facts and causes of behavior assuming that facts gathered do not change, while qualitative methods are concerned with the changing, dynamic nature of reality.
3. When quantitative approaches are used, the situation is structured by identifying and isolating specific variables for measurement

and by employing specific measurement tools and methods to collect information on these variables. In contrast, qualitative approaches attempt to gain a complete or holistic view of what is being measured by using a wide array of data including documents, records, photographs, observations, interviews, case histories, and even quantitative data.

4. Usually highly structured procedures, designed to verify or disprove predetermined hypotheses, are employed with quantitative approaches. Flexibility is kept to a minimum in an attempt to minimize bias. Procedures used with qualitative approaches, on the other hand, are usually flexible, exploratory, and discovery-oriented.
5. Quantitative approaches yield objective data that are typically expressed in numbers, while qualitative approaches focus on subjective data that are typically expressed or reported through language.
6. Quantitative data are usually collected under controlled conditions, while qualitative data are usually collected within the context of their natural occurrence.
7. In both approaches, reliability and validity are valued. In the quantitative approach, there is a heavy emphasis on reliability, that is, data that are consistent, stable, and replicable. Qualitative approaches, while recognizing that reliability is a necessary prerequisite for validity, tend to concentrate on validity, that is, data that are representative of a true and full picture of the phenomenon that is investigated (pp. 130–131).

Over time, more researchers have come to value using a mixed method approach and have begun to recognize the value of integrating qualitative and quantitative approaches within the context of a given study. It should be noted that the integration of qualitative and quantitative approaches is not simply mixing methods, but rather requires one to assume that the two approaches are complementary and that the primacy of the paradigmatic assumptions underlying one or the other approach can be eliminated as unproductive

(Haase & Myers, 1988). *Triangulation*, discussed in more detail in Chapter 25, is one methodological strategy for combining qualitative and quantitative approaches. In triangulation, multiple data sources, collection techniques, theories, and investigators are employed to assess the phenomenon of interest (Fielding & Fielding, 1986; Madey, 1982; Mitchell, 1986). Examples of the use of triangulation for combining qualitative and quantitative approaches can be found in Bretnauer, Ayres, and Knafl (1993), Floyd (1993), Corey (1993), Mason (1993), and Hendrickson (2008).

In addition to being categorized as norm-referenced or criterion-referenced, qualitative or quantitative, measuring tools and methods may be classified by (1) what they seek to measure, (2) the manner in which responses are obtained and scored, (3) the type of subject performance they seek to measure, or (4) who constructs them.

## What Is Measured

In nursing and health care research, there is usually interest in measuring cognition, affect, and psychomotor skills and/or physical functioning. *Cognitive* measures assess the subject's knowledge or achievement in a specific content area. Indicators of cognitive behavior usually are obtained as follows:

1. Achievement tests (objective and essay) that measure the extent to which cognitive objectives have been attained.
2. Self-evaluation measures designed to determine subjects' perceptions of the extent to which cognitive objectives have been met.
3. Rating scales and checklists for judging the specific attributes of products produced in conjunction with or as a result of an experience.
4. Sentence-completion exercises designed to categorize the types of responses and enumerate their frequencies relative to specific criteria.
5. Interviews to determine the frequencies and levels of satisfactory responses to formal and informal questions raised in a face-to-face setting.

6. Peer utilization surveys to ascertain the frequency of selection or assignment to leadership or resource roles.
7. Questionnaires employed to determine the frequency of responses to items in an objective format or number of responses to categorized dimensions developed from the content analysis of answers to open-ended questions.
8. Anecdotal records and critical incidents to ascertain the frequency of behaviors judged to be highly desirable or undesirable.
9. Review of records, reports, and other written materials (e.g., articles, autobiographical data, awards, citations, honors) to determine the numbers and types of accomplishments of subjects.

The number of cognitive measures employed far exceed the number of other types of measures. Specifically, written multiple-choice tests are the most often used, perhaps because they are the most objective of the various cognitive measures and the most reliable, and because they have the greatest utility in measuring all types of knowledge. Multiple-choice tests are further discussed in Chapters 4 and 17. Examples of cognitive measures can be found in Smith (1991), Grant et al. (1999), Story (2001), Arnold (2001), Tiro, Meissner, Korbin, and Chollete (2007), Boom, Nelson, Laufman, Kohrt, and Kozinetz (2007), Rondahl (2009), and design of mail and Internet surveys are discussed in Dillman (2007). It should be noted that cognitive measures are not limited to paper and pencil tests and that a variety of other approaches exist including computer-based testing, simulations that are discussed in Chapter 10 and Internet data collection discussed in Chapter 11.

*Affective* measures seek to determine interests, values, and attitudes. *Interests* are conceptualized as preferences for particular activities. Examples of statements relating to interests are:

- I prefer community-based nursing practice to practice in the hospital setting.
- I like to work with student nurses as they give care to patients.
- I prefer teaching responsibilities to administrative responsibilities.

- I would enjoy having one day a week to devote to giving direct care to patients in addition to my teaching responsibilities.

*Values* concern preferences for life goals and ways of life, in contrast to interests, which concern preferences for particular activities. Examples of statements relating to values are:

- I consider it important to have people respect nursing as a profession.
- A nurse's duty to her patient comes before duty to the community.
- Service to others is more important to me than personal ambition.
- I would rather be a teacher than an administrator.

*Attitudes* concern feelings about particular social objects, that is, physical objects, types of people, particular persons, or social institutions. Examples of statements relating to attitudes are:

- Nursing as a profession is a constructive force in determining health policy today.
- Continuing education for nurses should be mandatory for relicensing.
- Humanistic care is a right of all patients.
- All nurses should be patient advocates.

The feature that distinguishes attitudes from interests and values is that attitudes always concern a particular target or object. In contrast, interests and values concern numerous activities: specific activities in measures of interest and very broad categories of activities in measures of value. It is extremely difficult to preserve the conceptual differences among interests, values, and attitudes when actually constructing measures of affect. Thus, for the purpose of rendering them measurable, they are all subsumed under the rubric of *acquired behavioral dispositions* (Campbell, 1963) and are defined as tendencies to respond in a consistent manner to a certain category of stimuli. For example, when patients are asked to respond to a questionnaire to indicate their satisfaction with the quality of care received, one is interested in measuring their tendency to consistently respond that they are satisfied or dissatisfied, given a set of questions that ask them about

the care they received (the stimuli). Examples of the use of affective measures can be found in Grice, Picton, and Deakin (2003); Denny-Smith, Bairan, and Page (2005); Mackler, Wilkerson, and Cinti (2007); Gerend and Maglorie (2008).

*Self-report measures* are the most direct approach to the determination of affect. In this type of measure subjects are asked directly what their attitudes, interests, or values are. For example, subjects might be given a list of favorable and unfavorable statements regarding antagonistic patients and asked to agree or disagree with each. Such a self-report inventory is referred to as an attitude scale. Other indicators of affective behaviors include but are not limited to:

1. Sentence-completion exercises designed to obtain ratings of the psychological appropriateness of an individual's responses relative to specific criteria.
2. Interviews.
3. Questionnaires.
4. Semantic differential, Q-sort, and other self-concept perception devices.
5. Physiologic measures.
6. Projective techniques, for example, role playing or picture interpretation.
7. Observational techniques and behavioral tests, including measures of congruence between what is reported and how an individual actually behaves in a specific situation.
8. Anecdotal records and critical incidents.

Examples of self-report measures are included in Mackler et al. (2007) and Gerend and Maglorie (2008).

From the empirical evidence concerning the validity of different approaches, it appears that self-report offers the most valid approach currently available. For this reason, at present, most measures of affect are based on self-report and usually employ one of two types of scales: a summated rating scale or a semantic differential scale. A *scale* is a measuring tool or method composed of:

1. A stem, which is a statement relating to attitudes or an attitudinal object to be rated by the respondent.

Indicate your degree of agreement with the following statement:

**STEM:** Noncompliance on the part of the patient indicates a need for additional attention to be directed toward the quality of care received by the patient.

**SCALE**

**STEPS:** / / / / / / / / / / / / / / / /  
1 2 3 4 5 6 7 8 9 10 11 12

**ANCHORS:**

Completely  
Disagree

Completely  
Agree

**FIGURE 1.2** The components of a scale.

2. A series of scale steps.
3. Anchors that define the scale steps.

An example of a scale can be found in Chiravalle and McCaffrey (2005).

Figure 1.2 presents examples of the components of a scale.

There are different types of anchors that can be employed: numbers, percentages, degrees of agreement/disagreement, adjectives (e.g., worthless/valuable), actual behavior, and products (e.g., samples of nursing care plans to be rated 1 to 6). Usually numerical anchors are preferred for the following reasons: if the meaning of each step on the scale is specified at the beginning of the rating form, as is usually the case, numbers provide an effective means of coordinating those definitions with rating scales; numbers on scales constantly remind subjects of the meanings of scale steps; numbers facilitate the analysis of data, for example, inputting ratings for computer analysis (Nunnally, 1967; Nunnally & Bernstein, 1994).

### Summated Rating Scale

A summated rating scale contains a set of scales, all of which are considered approximately equal in attitude or value loading. The subjects respond with varying degrees of intensity on a scale ranging between extremes such as agree/disagree, like/dislike, or accept/reject. The scores of all scales in the set are summed or summed and averaged to yield an individual's attitude

score. An example of a summated rating scale is given in Figure 1.3.

Summated rating scales are easy to construct, are usually reliable, and are flexible in that they may be adapted for the measurement of many different kinds of attitudes. Nunnally (1967) and Nunnally and Bernstein (1994) suggest that the reliability of summated scales is a direct function of the number of items. When there are a reasonable number of items (e.g., 20) on the scale, fewer scale steps for individual scales are required for a high degree of reliability. When there are fewer items, more scale steps for individual scales are required for reliability. In most cases, 10 to 15 items using 5 or 6 steps are sufficient. Individual scales on summated attitude scales tend to correlate substantially with each other, because it is fairly easy for the constructor to devise items that obviously relate to each other and for subjects to see the common core of meaning in the items. Additional information regarding summated attitude scales can be found in Edwards (1957), Shaw and Wright (1967), Nunnally (1967), or Nunnally and Bernstein (1994).

An example of a summated rating scale can be found in Lehoux and colleagues (2006).

### Semantic Differential Scales

The semantic differential is a method for measuring the meaning of concepts that was developed by Osgood, Suci, and Tannenbaum (1957). The semantic differential has three components: (1) the



Rate the following concept in terms of how you feel about it at this point in time:

Noncomplying Patient

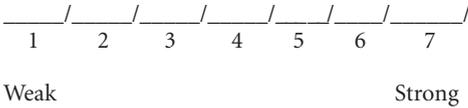
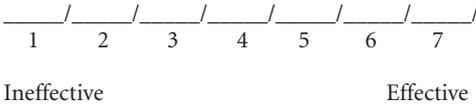


FIGURE 1.4 Example of a semantic differential scale.

TABLE 1.1 Frequently Employed Anchors for Semantic Differential Factors

	Factor		
	Evaluation	Potency	Activity
BIPOLAR ADJECTIVES	good ..... bad	strong ..... weak	active ..... passive
	fair ..... unfair	large ..... small	quick ..... slow
	positive ..... negative	severe ..... lenient	tense ..... relaxed
	honest ..... dishonest	hard ..... soft	sharp ..... dull
	successful ..... unsuccessful		
	valuable ..... worthless		

and Freeman (2001), Kelly (2001), Chiravalle and McCaffrey (2005).

*Psychomotor measures* seek to assess subjects' skill, that is, their ability to perform specific tasks or carry out specific procedures, techniques, and the like. An important consideration in the measurement of psychomotor objectives involves the manner in which the skills and materials or objects to be manipulated or coordinated are specified. Specifically, criteria for the successful manipulation of an object must be clearly and unambiguously stated at the time when objectives are made explicit. Task-analysis procedures (Gagne, 1962) are often used to accomplish this.

The most viable approach to the measurement of psychomotor skills at this time is the observation method combined with a performance

checklist or rating scale. The observation method always involves some interaction between subject and observer in which the observer has an opportunity to watch the subject perform. Checklists and rating scales are often employed to record in a systematic manner behaviors and events occurring during the observation period. A checklist is most often used to note the presence or absence or the frequency of occurrence of the specified behavior, conditions, characteristics, or events. Rating scales are employed when it is desirable that the observer rate the behavior in terms of points along a continuum. Ratings may occur at predetermined intervals throughout the observation session or at the end of the observation to summarize what occurred during the observation period.

The conceptual definition of the phenomenon to be measured forms the basis for how and where the observation should occur. Qualitative conceptual perspectives generally lead to free and unstructured observations in naturalistic settings, while quantitative conceptualizations more often lead to more structured observations using guidelines and trained observers. The work of Cohn, Matias, Tronick, Connell, and Lyons (1986) provides an example of a study combining both structured, standardized observation methods and naturalistic unstructured observations of depressed mothers and their infants. The reader will find an interesting comparison and contrast between the two approaches for collecting observational data.

*Unstructured and semistructured* observations involve the collection of descriptive information that is generally analyzed in a qualitative manner. When *structured* observation is employed, it is often necessary that the nurse prepare an observation guide to structure the observation and train the observer in its use. This guide increases the probability that the crucial behaviors of concern will be considered, which increases the reliability and validity of the method.

When possible, the guide should specify when an observation begins and ends as well as what behaviors are to be observed. Frequently, time is the vehicle for accomplishing this purpose. For example, it might be specified that observation of medical students' history taking begin when the student enters the room and continue for the first 5 minutes of the student-patient interaction. Using a more elaborate scheme, observations will begin when the student enters the room, and will continue for 2 minutes; then observers will rest for the next 2 minutes, rate for 2 minutes, rest for 2 minutes, and so forth, until the student leaves the room and the encounter ends. No matter how structured the observation or how well trained or competent the observer, observation techniques, in order to be sound, require more than one observer. This provides an estimate of the accuracy or reliability of the observations and provides a basis for determining the degree of confidence to be placed in the data.

Three factors must be considered in the discussion of observational techniques: (1) inter-

action between respondents and observers, (2) whether respondents know they are being observed, and (3) whether respondents know when they will be observed. Observation is difficult because watching a situation often changes it so that the observers are no longer certain of what they are observing. This implies that a basic criterion for evaluating studies in which observation is used is the extent to which the situation observed was natural. Observations of a subject's psychomotor skills should be accomplished with as little effect as possible on the natural situation in which the skills are normally performed. Webb et al. (1966) have published a useful book full of suggestions about how measures can be collected as unobtrusively as possible.

It is necessary to weigh the value of collecting observational data over time as opposed to collecting information at one isolated point in time. Observational data collected at one time are subject to more errors of measurement and, hence, lower reliability and validity than observational data collected at multiple times. When subjects' psychomotor performance is of interest, the concern is usually with how they perform most of the time, or typically; that is, patterns of performance or consistency in performance over time becomes important. When observational data are collected at one point in time, there is greater probability that the results of the measurement will reflect more of the conditions surrounding that isolated point in time than the true abilities of the subjects to perform the tasks or behaviors. Hence, whenever possible, measures of performance should occur at more than one point in time.

Observational techniques may be direct or indirect. In *direct observation* the observer evaluates psychomotor performance by simply watching the subject perform. A limitation of this approach stems from the fact that it is both time-consuming and expensive. It is, however, an excellent technique for the assessment of behavior in conjunction with clinical performance, especially when the concern is with dynamic or process variables. Similarly, a unique strength of the observation method results from the fact that if the observer wishes to learn how a subject functions under the pressure of supervision,

there is no substitute for direct observation that is known and scheduled.

*Indirect observation* methods include motion picture, television, videotaping, and other devices for recording subjects' activities. The value of indirect techniques results from the opportunities they afford subjects to become involved in the evaluation of their performance as the recording is viewed jointly by the observer and respondents. Indirect observations are limited in that they are not sensitive to the tone, mood, or affect of the situation. Another limitation is that mechanical devices selectively record, depending on their placement, where they are aimed by the operator and, hence, the total situation may be missed. This limitation can be turned into an advantage, however, if multiple devices are used to record all that happens before them. Examples of psychomotor measures can be found in Bujak, McMillan, Dwyer, and Hazelton (1991); DeMattes, et al. (1993); Mason and Redeker (1993); Finke et al. (2001); Mims (2001); Kostopoulos (2001); Gilbert, Temby, and Rogers (2004); Philpin (2006); Williams (2006); and Cricco-Lizza (2006). Observational methods are discussed in more detail in Chapter 12.

*Physiologic measures* seek to quantify the level of functioning of living beings. Indicators of physiologic functioning include but are not limited to:

1. Blood pressure readings.
2. Temperature readings.
3. Respiratory measures.
4. Metabolic readings.
5. Diabetic and other screening devices.
6. Readings from cardiac and other monitoring instruments.
7. ECG and EEG readings.
8. Results of blood tests and analyses.
9. Measures of height and weight.

Physical functioning can often be measured by a scientific instrument, and the results of physiologic measures usually are expressed as a quantitative scale that can be graded into finely distinguished numerical values. For example, the variable diastolic blood pressure is measured using a scientific instrument referred to as a sphygmomanometer. Its scale is in a quantitative

form ranging from 0 to 300, providing a total of 300 different continuous scale points or values to which a subject can be assigned and which differentiate among the various degrees of the variable possessed by the subjects measured. Thus, on the basis of blood pressure readings, one can state that a subject with a diastolic pressure of 100 is 20 points higher than one with a diastolic pressure of 80. This 20-point difference is significant in comparing the physical status of two patients.

Well-designed and implemented physiologic measures are among the most precise methods one can employ; they yield data measured at the interval or ratio level of measurement, allow a wide range of statistical procedures to be employed in their analysis, and tend to produce results that demonstrate a high degree of reliability and validity. Examples of physiologic measures can be found in Heidenreich and Giuffrè (1990), Bridges and Woods (1993), Partridge and Hughes (2007). The physiologic approach to measurement is discussed in detail in Chapter 20.

## How Responses Are Obtained and Scored

The distinction to be considered is whether a measure is objective or subjective. It should be noted that a given method or technique is generally viewed as more or less objective or subjective; that is, one may think in terms of a continuum anchored by the terms objective and subjective, and then place a given method on the continuum, depending on whether it possesses characteristics more like those of an objective or subjective measure.

*Objective measures* contain items that allow subjects little if any latitude in constructing their responses and spell out criteria for scoring so clearly that scores can be assigned either by individuals who know nothing of the content or by mechanical means. Multiple-choice questions and physiologic measures are examples of the most objective methods that can be employed.

*Subjective measures* allow respondents considerable latitude in constructing their responses. In addition, the probability that different scorers may apply different criteria is greater. Examples of subjective measures are the essay test,

open-ended interview questions, case studies, and nursing care plans. The essay question is a method requiring a response constructed by the subject, usually in the form of one or more sentences. The nature of the response is such that (1) no single answer or pattern of answers can be listed as correct, and (2) the quality of the response can be judged only subjectively by one skilled or informed in the subject (Stalnaker, 1951). Thus, significant features of the essay method are (1) the freedom of response allowed the respondents, (2) the fact that no single answer can be identified as correct or complete, and (3) responses must be scored by experts who themselves usually cannot classify a response as categorically right or wrong. Essay questions may require subjects to express their own thoughts on an issue of interest to the profession, outline a research design for investigating a research question, derive a mathematical proof, or explain the nature of some nursing phenomenon. Items may require only a brief response or may demand an extensive exposition.

Advocates of the essay approach argue that an important characteristic of individuals is their ability to interact effectively with other individuals in the realm of ideas. The basic tool of interaction is language, and successful individuals are those who can react appropriately to questions or problems in their field as they encounter them. It is not enough, they contend, to be able to recognize a correct fact when it is presented or to discriminate among alternatives posed by others. Successful individuals are the masters of their collection of ideas and are able to cite evidence to support a position and contribute to the advancement of ideas and constructs within their field. The only way to assess the extent to which individuals have mastered a field is to present them with questions or problems in the field and assess how they perform. Hence, they argue, the essay format provides an avenue for assessing scholarly and/or professional performances better than other available methods (Coffman, 1971).

Even so, because of their subjective nature, essay questions have inherent limitations that must be recognized and minimized if sound measurement is to result from their use. The limitations of essays and other subjective measures fall into

two general categories: (1) problems related to the difficulty in achieving consistency in scoring responses, and (2) problems associated with the sampling of content. Empirical evidence regarding the reliability of subjective measures suggests that different raters tend to assign different scores to the same response, a single rater tends to assign different scores to the same response on different occasions, and the differences tend to increase as the measure permits greater freedom of response (Finlayson, 1951; Hartog & Rhodes, 1936; Noyes, 1963; Pearson, 1955; Vernon & Millican, 1954). Different raters may differ as a result of a number of factors, including the severity of their standards, the extent to which they distribute scores throughout the score scale, and real differences in the criteria they are applying.

Basic to problems associated with the sampling of content is the notion that each sample unit should be independent and equally likely to be chosen in the sample. In general, the greater the number of different questions, the higher the reliability of the score. The compromise to be made, however, is between the desire to increase the adequacy of the sample of content by asking many different questions and the desire to ask questions that probe deeply the subjects' understanding. Additional information regarding essay items is presented in Stalnaker (1951) and Coffman (1971). Examples of structured, unstructured, and semistructured interviews can be found in Irwin and Johnson (2005); Cricco-Lizza (2006); and Clarke, Booth, Velikova, and Hewison (2006), respectively. Other subjective measures are described in more detail in Part IV.

## Type of Performance Measured

When performance is of interest, one may seek to measure typical performance or maximum performance. If the interest is in assessing subjects as they do their best (produce their highest quality work), then a *maximum performance* measure is appropriate. Such measures are indices of cognition that generally measure a set of skills a subject possesses but that differ among themselves in the specificity of their focus and the use to which scores are put. Maximum performance measures of particular interest include

aptitude measures, achievement measures, and diagnostic measures.

*Aptitude* tests are specific measures of capacity for success and tend to focus on various general aspects of human ability (e.g., mechanical aptitude, artistic aptitude). They are often used as predictors of performance in special fields.

*Achievement* measures are tests of particular skills and knowledge and are more specific than aptitude tests. They usually sample a wide range of skills and are constructed by nurses and other health professionals for their own use. Commercially produced achievement measures are also available in many different content areas. *Diagnostic* tests are even more specific in their focus than achievement measures, although this need not always be the case. They focus on specific skills and often employ multiple measures of particular skills. Their intent is to pinpoint specific weaknesses that might not be apparent otherwise. Once specific deficiencies are identified and remediation has taken place, one might predict that achievement, which is assumed to be dependent on these more specific skills, will improve.

If information about subjects' typical behavior (i.e., what they usually do or would do) is of interest, it is appropriate to use a *typical performance* measure. These are measures of affective behavior and usually attempt to have respondents describe the way they typically perceive themselves or their behavior. Typical performance measures usually ask the subjects for scaled responses, forced-choice responses, or criterion-keyed responses. Exhibit 1.1 presents examples of each of these types of responses.

## Who Constructs Measures

*Standardized* measures are developed by specialists for wide use. Their content is set, the directions for administration (often including time limits) are clearly described, and the scoring procedure to be used is completely prescribed. Information on norms concerning scores is generally available. Examples of standardized measures employed to assess the outcomes of nursing education programs are presented in *Educational Outcomes: Assessment of Quality—A Prototype for Student Outcome Measurement in*

*Nursing Programs* (Waltz, 1988). In Chapter 7, standardized approaches to measurement are discussed in detail.

*Informal tools and methods* are typically constructed by nurses and other health professionals for their own use. They are not content-constrained; that is, the user is free to define the content as well as administration procedures and scoring. Norms may be available for local groups but more often are not available for any group.

In summary, the measurement framework employed in a given situation will have important implications for instrument development and for what can be done with and on the basis of the resulting information. Thus, it is important to clarify at the outset the type of measurement that will yield data appropriate for the types of questions and/or hypotheses one seeks to answer. In Chapters 2 through 23, attention is focused on instrument development and testing in both the norm-referenced and criterion-referenced cases. In Chapters 24 through 26, measurement issues and important considerations to be made in using the types of measures presented in this section are addressed.

## RELIABILITY AND VALIDITY OF MEASURES

As indicated in the foregoing sections, reliability and validity are essential characteristics of any measuring tool or method. Factors that may affect the degree of consistency obtained for a given measure (reliability) are (1) the manner in which the measure is scored; (2) characteristics of the measure itself; (3) the physical and/or emotional state of the individual at measurement time; and (4) properties of the situation in which the measure is administered (e.g., the amount of noise, lighting conditions, temperature of the room).

Strictly speaking, one validates not the measurement tool or method but rather some use to which the measure is put. For example, an instrument designed to select participants who would benefit from a primary care fellowship experience must be valid for that purpose, but it would not necessarily be valid for other purposes such as measuring how well participants master

**EXHIBIT I.1 Sample Responses of Typical Performance Measures****Scaled response**

When a scaled response is employed, the respondent indicates on a scale what his/her rating or answer is to a question posed. For example:

Do findings from this research study provide information that will be meaningful to you in your clinical practice? Please rate.

not at all	very little	somewhat	enough	a lot
1	2	3	4	5

**Forced-choice response**

With a forced-choice response item, the respondent is asked to choose between 2 or more different alternatives, all of which may be appealing responses. The point is that one particular response is most appealing to the subject. For example:

A program of ongoing evaluation and implementation of research findings that may serve as a basis for practice does not exist in the agency with which you are affiliated. You are aware of the need to develop such a program. You would prefer to have this need met by:

1. Referring the task to someone else.
2. Supporting activities of the professional nursing organizations that are seeking to increase involvement in implementing evidence-based practice in all clinical settings.
3. Supporting a policy change in the agency responsible for care.
4. Serving as a resource person to staff by providing them with knowledge and materials to enable them to develop such a program.
5. Becoming a member of a committee of practitioners who are developing and testing a pilot program in conjunction with the patients for whom they deliver care.
6. Initiating the idea of such a program by implementing evidence-based practice for patients on your unit and sharing with staff the various approaches you employ in evaluating and implementing research findings in your practice.

**Criterion-keyed response**

Criterion-keyed responses depend on information previously obtained about how certain groups answered the items. If a subject's score looks like those of members of a predefined group, he/she is classified as a member of that group. The assumption is that the criterion for membership in a specific group is having a set of responses on the measure that looks like those from the predefined group. For example, the Minnesota Multiphasic Personality Inventory (MMPI) was originally used with hospitalized psychiatric patients and normal (i.e., nonhospitalized subjects) to construct a criterion-keyed set of questions that had some value as predictors of mental stability, that is, if a specific item was responded to differently by the two groups, it was included on the measure.

objectives at the completion of the fellowship experience.

Both reliability and validity are matters of degree rather than all-or-none properties. Measures should be assessed each time they are used to see if they are behaving as planned. New evidence may suggest modifications in an existing measure or the development of a new and better approach to measuring the attribute in question. Reliability is a necessary prerequisite for validity; that is, if a measure does not assign scores consistently, it cannot be useful for the purpose for which it is intended. Reliability is not, however, a sufficient condition for validity; that is, because a measure consistently measures a phenomenon does not ensure that it measures the phenomenon of interest.

As stated earlier, the determination of the reliability and validity of a specific tool or method will differ depending on whether it is norm-referenced or criterion-referenced. Specific techniques for determining reliability and validity in each case are discussed in Chapters 5 and 6. In either case, the reliability and validity of the measurement process itself is increased when multiple measures of the same thing are employed; that is, more than one type of instrumentation is used to answer a given question. Similarly, reliability and validity increase when the answer to a given measurement concern is elicited by collecting data from a number of different sources using the same measurement tool or method.

Evidence for reliability and validity of a tool or method is accrued over time, so although

one cannot rely on “old” reliability and validity evidence, it is important to seek information regarding reliability and validity testing results in previous studies. Often difficulties are encountered in locating information regarding reliability and validity evidence for existing tools. An article by Strickland (2006) presents strategies for searching for evidence of reliability and validity for an existing instrument that may be useful in this regard.

## REFERENCES

- Adam, P. F., & Freeman, L. H. (2001). Assertive behavior inventory tools. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 245–250). New York: Springer Publishing.
- Aitken, L. M., Pelter, M. M., Carlson, B., Marshall, A. P., Cross, R., McKinley, S., et al. (2008). Effective strategies for implementing a multicenter international clinical trial. *Journal of Nursing Scholarship*, 40(2), 101–108.
- Allegood, M. R., & Marriner-Tomey, A. (2002). *Nursing theory: Utilization and application*. St. Louis: Mosby Year Books.
- American College of Cardiology and the American Heart Association. (2007). *Guidelines for the Management of Patients with Chronic Stable Angina*. Retrieved June 20, 2008, from <http://circ.ahajournals.org/cgi/content/full/116/23/2762>
- Armitage, C. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology*, 40, 471–499.
- Arnold, J. M. (2001). Diagnostic reasoning simulations and instruments. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 3–21). New York: Springer Publishing.
- Arwood, E., & Kaakinen, J. (2009). SIMulation based on language and learning (SIMBALL): The model. *International Journal of Nursing Education Scholarship*, 6(1), Article 9. Retrieved August 28, 2009, from <http://www.bepress.com/ijnes/vol6/iss1/art9>
- Avci, I. A., & Kurt, H. (2008). Health beliefs and mammography rates of Turkish women living in rural areas. *Journal of Nursing Scholarship*, 40(2), 170–175.
- Bergstrom, N. (1989). Nursing diagnosis research: Inductive and deductive approaches. In R. M. Carroll-Johnson (Ed.), *Proceedings of Invitational Conference on Research Methods for Validating Nursing Diagnosis*. Palm Springs, CA: North American Nursing Diagnosis Association and Case Western University.
- Boom, J. A., Nelson, C. S., Laufman, L. E., Kohrt, A. E., and Kozinetz, C. A. (2007). Improvement in provider immunization knowledge and behaviors following a peer education intervention. *Clinical Pediatrics*, 46(8), 706–717.
- Bramadat, I. J., & Driedger, M. (1993). Satisfaction with childbirth: Theories and methods of measurement. *Birth: Issues in Prenatal Care and Education*, 20(1), 22–29.
- Breitnauer, B. J., Ayres, I., & Knafl, K. A. (1993). Triangulation in qualitative research: Evaluation of completeness and confirmation purposes. *Image: Journal of Nursing Scholarship*, 25(3), 237–243.
- Bridges, E. J., & Woods, S. J. (1993). Pulmonary artery pressure measurement: State of the art. *Heart and Lung: Journal of Critical Care*, 22(2), 99–111.
- Broome, M. E., & Mobley, T. (2003). The child medical fears scale. In O. L. Strickland & C. Dilorio (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 2, pp. 196–206). New York: Springer Publishing.
- Bujak, L., McMillan, M., Dwyer, J., & Hazelton, M. (1991). Assessing comprehensive nursing performance: The objective structural clinical assessment. *Nurse Education Today*, 11(3), 179–184.
- Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 6, pp. 94–172). New York: McGraw-Hill.
- Campbell, D. T., & Fiske, O. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105.
- Chenitz, W. C., & Swanson, J. M. (1986). *From practice to grounded theory: Qualitative research in nursing*. Menlo Park, CA: Addison-Wesley.

- Chiravalle, P., & McCaffrey, R. (2005). Alternative therapy applications for postoperative nausea and vomiting. *Holistic Nursing Practice, 19*(5), 207–210.
- Clarke, S., Booth, L., Velikova, G., & Hewison, J. (2006). Social support: Gender differences in cancer patients in the United Kingdom. *Cancer Nursing, 29*(1), 66–72.
- Coffman, E. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 271–302). Washington, DC: American Council on Education.
- Cohn, J. F., Matias, R., Tronick, E. Z., Connell, D., & Lyons, R. K. (1986). Face-to-face interactions of depressed mothers and their infants. In E. Z. Tronick & T. Fields (Eds.), *New Directions for Child Development, 34*, 31–45.
- Corey, J. W. (1993). Linking qualitative and quantitative methods: Integrating cultural factors into public health. *Qualitative Health Research, 3*(3), 298–318.
- Cousins, S. O. (1997). Validity and reliability of self-reported health of persons aged 70 and older. *Health Care for Women International, 18*(2), 165–174.
- Cricco-Lizza, R. (2006). Black non-Hispanic mothers' perceptions about the promotion of infant-feeding methods by nurses and physicians. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 35*(2), 173–180.
- Czar, M. I., & Engler, M. M. (1997). Perceived learning needs of patients with coronary artery disease using a questionnaire assessment tool. *Heart & Lung, 26*(2), 109–117.
- DeMattes, O., Law, M., Russell, D., Pollock, N., Rosenbaum, P., & Walter, S. (1993). The reliability and validity of the quality of upper extremity skills test. *Physical and Occupational Therapy in Pediatrics, 13*(2), 1–18.
- Denny-Smith, T., Bairan, A., & Page, M. C. (2005). A survey of female nursing students' knowledge, health beliefs, perceptions of risk, and risk behaviors regarding HPV and cervical cancer. *Journal of the American Academy of Nurse Practitioners, 18*(1), 62–69.
- Diekelmann, N. L. (1992). Learning as listening: A Heideggerian hermeneutical analysis of the lived experiences of students and teachers in nursing. *Advances in Nursing Science, 14*(3), 72–83.
- Dillman, D. (2007). *Mail and internet surveys: The Tailored Design Method* (2nd ed.). Hoboken, NJ: Wiley.
- Duffy, M. E. (1987). Methodological triangulation: A vehicle for merging quantitative and qualitative research methods. *Image: Journal of Nursing Scholarship, 19*(3), 130–133.
- Edwards, A. (1957). *Techniques of attitude scale construction*. New York: Appleton-Century Crofts.
- Ellenbecker, C. H., Porell, F. W., Samia, L., Byleckie, J. J., & Milburn, M. (2008). Predictors of home healthcare nurse retention. *Journal of Nursing Scholarship, 40*(2), 151–160.
- Escoffery, C., Glanz, K., Hall, D., & Elliott, T. (2009). A multi-method process evaluation for a skin cancer prevention diffusion trial. *Evaluation & the Health Professions, 32*(2), 194–203.
- Evans, S., Tsao, J., Sternlieb, B., & Zeltzer, L. K. (2009). Using the Biopsychosocial Model to understand the health benefits of yoga. *Journal of Complementary and Integrative Medicine, 6*(1), Article 15. Retrieved August 28, 2009, from <http://www.bepress.com/jcim/vol6/iss1/15>
- Fielding, N., & Fielding, J. (1986). *Linking data: Qualitative research methods, series 4*. Beverly Hills, CA: Sage Publications.
- Filstead, W. J. (1974). *Qualitative methodology: First-hand involvement with the social world*. Chicago: Markham.
- Finke, L., Messmer, P., Sprock, M., Gilman, B., Weiner, E., & Emerson, L. A. (2001). Measuring RN students' clinical skills via computer. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 194–198). New York: Springer Publishing.
- Finlayson, D. S. (1951). The reliability of the marking of essays. *British Journal of Educational Psychology, 21*, 126–134.
- Floyd, J. R. (1993). The use of across-method triangulation in the study of sleep concerns in healthy older adults. *Advances in Nursing Science, 16*(2), 70–80.
- Fry, S. T. (2001). Justification of moral judgment and action tool. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 276–281). New York: Springer Publishing.

- Gagne, R. M. (1962). The acquisition of knowledge. *Psychological Review*, 69, 255–476.
- George, J. B. (2002). *Nursing theories* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Gerend, M. A., and Maglorie, Z. F. (2008). Awareness, knowledge and beliefs about human papillomavirus in a racially diverse sample of young adults. *Journal of Adolescent Health*, 42, 237–242.
- Gilbert, L., Temby, J., & Rogers, S. (2004). *Evaluating a teen STD prevention web site*. Retrieved July 13, 2008, from <http://www.ncbi.nlm.nih.gov.proxy-hs.researchport.umd.edu/sites/entrez>
- Girot, E. A. (1993). Assessment of competence in clinical practice: A phenomenological approach. *Journal of Advanced Nursing*, 18(1), 114–119.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine Publishing.
- Grant, E., Turner, K., Daugherty, S., Li, T., Eckenfels, E., Baier, C., et al. (1999). Development of a survey of asthma knowledge, attitudes, and perceptions: The Chicago Community Asthma Survey. *Chest*, 116(4), 178–183.
- Grice, A. S., Picton, P., Deakin, C. D. S. (2003). Study examining attitudes of staff, patients and relatives in adult intensive care units. *British Journal of Anaesthesia*, 91(6), 820–824.
- Haase, J. E., & Myers, S. T. (1988). Reconciling paradigm assumptions of qualitative and quantitative research. *Western Journal of Nursing Research*, 10(2), 128–137.
- Hartog, P., & Rhodes, E. C. (1936). *The marks of examiners*. New York: Macmillan.
- Heidenreich, L., & Giuffre, M. (1990). Postoperative temperature measurement. *Nursing Research*, 39(3), 154–155.
- Hendrickson, S. G. (2008). Maternal worries, home safety behaviors, and perceived difficulties. *Journal of Nursing Scholarship*, 40(2), 137–143.
- Hinds, P. S., & Young, K. J. (1987). A triangulation of methods and paradigms to study nurse-given wellness care. *Nursing Research*, 36(3), 195–198.
- Irwin, L., & Johnson, J. (2005). Interviewing young children: Explicating our practices and dilemmas. *Qualitative Health Research*, 15(6), 821–831.
- Kaji, A., Koenig, K., & Bey, T. (2006). Surge capacity for healthcare systems: A conceptual framework. *Academic Emergency Medicine*, 13, 1157–1159.
- Kelly, K. (2001). Nursing activity scale. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 253–266). New York: Springer Publishing.
- Kempen, G., Miedema, I., Ormel, J., & Molenaar, W. (1996). The assessment of disability with the Groningen Activity Restriction Scale: Conceptual framework and psychometric properties. *Social Science and Medicine*, 43(11), 1601–1610.
- Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- King, L. M. (2003). Assessment of functional abilities and goal attainment scales: A criterion-referenced measure. In O. L. Strickland & C. Dilorio (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 2, pp. 3–20). New York: Springer Publishing.
- Knafl, K. A. (1989). Concept development. In R. M. Carroll-Johnson (Ed.), *Proceedings of Invitational Conference on Research Methods for Validating Nursing Diagnosis*. Palm Springs, CA: North American Nursing Diagnosis Association and Case Western University.
- Kostopoulos, M. R. (2001). Performance appraisal tool. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 79–91). New York: Springer Publishing.
- Krogh, G., & Naden, D. (2008). A nursing-specific model of EPR documentation: Organizational and professional requirements. *Journal of Nursing Scholarship*, 40(1), 68–75.
- Landis, A. M., Parker, K. P., & Dunbar, S. B. (2009). Sleep, hunger, satiety, food cravings, and caloric intake in adolescents. *Journal of Nursing Scholarship*, 41(2), 115–123.
- Lee, B., Chaboyer, W., & Wallis, M. (2008). Predictors of health-related quality of life 3 months after traumatic injury. *Journal of Nursing Scholarship*, 40(1), 83–90.
- Lehoux, P., Richard, L., Pineault, R., & Saint-Arnaud, J. (2006). Delivery of high-tech home care by hospital-based nursing units in Quebec. *Canadian Journal of Nursing Leadership*, 19(1), 44–55.

- Leonard, M. K., & George, J. B. (1995). In J. B. George (Ed.), *Nursing theories: The base for professional nursing practice* (4th ed., pp. 145–164). East Norwalk, CT: Appleton & Lange.
- LoBiondo-Wood, G., & Haber, J. (1994). *Nursing research methods, critical appraisal and utilization* (3rd ed.). St. Louis: CV Mosby.
- Mackler, N., Wilkerson, W., and Cinti, S. (2007). Will first-responders show up for work during a pandemic? Lessons from a smallpox vaccination survey of paramedics. *Disaster Management & Response*, 5(2), 45–48.
- Madey, D. L. (1982). Benefits of qualitative and quantitative methods in program evaluation, with illustrations. *Educational Evaluation and Policy Analysis*, 4(2), 223–236.
- March, J., & Olsen, F. (2006). The logic of appropriateness. In M. Moran, M. Rein, & R. Goodin (Eds.), *The Oxford Handbook of Public Policy* (pp. 698–708). New York: Oxford University Press.
- Mason, D. J., & Redeker, N. (1993). Measurement of activity. *Nursing Research*, 42(2), 87–92.
- Mason, S. A. (1993). Employing quantitative and qualitative methods in one study. *British Journal of Nursing*, 2(17), 869–872.
- Mbweza, E., Norr, K. F., & McElmurry, B. (2008). Couple decision making and use of cultural scripts in Malawi. *Journal of Nursing Scholarship*, 40(1), 12–19.
- Melby, C. S., Dodgson, J. E., & Tarrant, M. (2008). The experiences of western expatriate nursing educators teaching in Eastern Asia. *Journal of Nursing Scholarship*, 40(2), 176–183.
- Meleis, A. L. (1999). *Theoretical nursing: Development and progress* (3rd ed.). New York: J. B. Lippincott.
- Mims, B. C. (2001). Clinical performance examination for critical care nurses. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed., Vol. 1, pp. 102–111). New York: Springer Publishing.
- Mitchell, E. S. (1986). Multiple triangulation: A methodology for nursing science. *Advances in Nursing Science*, 8(3), 10–26.
- Noyes, E. S. (1963). Essay and objective tests in English. *College Board Review*, 49, 7–10.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Olsen, J. (2007). *Understanding institutions and logics of appropriateness: Introductory Essay*. Retrieved July 13, 2008, from [www.arena.uio.no/publications/working-papers2007/papers/wp07\\_13.pdf](http://www.arena.uio.no/publications/working-papers2007/papers/wp07_13.pdf)
- Orem, D. E. (1995). *Nursing concepts of practice* (5th ed.). St. Louis: Mosby Year Books.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago: University of Illinois Press.
- Partridge, J., & Hughes, J. (2007). Genital human papillomavirus infection in men: Incidence and risk factors in a cohort of university students. *Journal of Infectious Disease*, 196, 1128–1136.
- Pearson, R. (1955). The test fails as an entrance examination. Should the general composition test be continued? *College Board Review*, 25, 209.
- Perez-Woods, R. C., Burns, B., & Chase, C. I. (1989). The anatomy of NCC certification examinations: The Neonatal Intensive Care Nursing Examination. *Neonatal Network Journal of Neonatal Nursing*, 7(6), 53–59.
- Philpin, S. (2006). “Handing over”: Transmission of information between nurses in an intensive therapy unit. *Nursing in Critical Care*, 11(2), 86–93.
- Polit, D., & Beck, C. (2008). *Nursing Research—Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott, Williams and Wilkins.
- Rempusheski, V. F., & O’Hara, C. (2005). Psychometric properties of the Grandparent Perceptions of Family Scale (GPFS). *Nursing Research*, 54(6), 419–427.
- Renpenning, K. M., & Taylor, S. G. (Eds.). (2003). *Self-care theory in nursing: Selected papers of Dorothea Orem*. New York: Springer Publishing.
- Rew, L., Grady, M., Whittaker, T. A., & Bowman, K. (2008). Interaction of duration of homelessness and gender on adolescent sexual health indicators. *Journal of Nursing Scholarship*, 40(2), 109–115.
- Rondahl, G. (2009). Students’ inadequate knowledge about lesbian, gay, bisexual and

- transgender persons. *International Journal of Nursing Education Scholarship*, 6(1), Article 11. Retrieved August 28, 2009, from <http://www.bepress.com/ijnes/vol6/iss1/art11>
- Rubin, H. R., Pronovost, P., & Diette, G. B. (2001). From a process of care to a measure: The development and testing of a quality indicator. *International Journal for Quality in Health Care*, 13(6), 489–496.
- Schneider, J. S., Barkauskas, V., & Keenan, G. (2008). Evaluating home health care nursing outcomes with OASIS and NOC. *Journal of Nursing Scholarship*, 40(1), 76–82.
- Shaw, R. E., & Wright, J. M. (1967). *Scales for the measurement of attitudes*. New York: McGraw-Hill.
- Smith, M. M. (1991). Coronary Heart Disease Knowledge Test: Developing a valid and reliable tool. *Nurse Practitioner: American Journal of Primary Health Care*, 16(4), 28, 31, 35–36.
- Snider, J. G., & Osgood, C. E. (Eds.). (1969). *Semantic differential technique*. Chicago: Aldine Publishing.
- Stacey, D., Higuchi, K. A., Menard, P., Davis, B., Graham, I., & O'Connor, A. M. (2009). Integrating patient decision support in an undergraduate nursing curriculum: An implementation project. *International Journal of Nursing Education Scholarship*, 6(1), Article 10. Retrieved August 28, 2009, from <http://www.bepress.com/ijnes/vol6/iss1/art10>
- Stainbeck, S., & Stainbeck, N. (1984). Broadening the research perspective in education. *Exceptional Children*, 50, 400–408.
- Stalnaker, J. M. (1951). The essay type of examination. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 495–530). Washington, DC: American Council on Education.
- Stein, R. F., & Green, E. J. (1970). The Graduate Record Examination as a predictive potential in the nursing major. *Nursing Research*, 19(1), 42–47.
- Story, D. K. (2001). Measuring clinical decision making using a clinical simulation film. In C. F. Waltz & L. S. Jenkins (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 1, pp. 102–111). New York: Springer Publishing.
- Strickland, O. (2006). In search of reliability and validity evidence for an existing instrument. *Journal of Nursing Measurement*, 14(1), 3–4.
- Tiro, J., Meissner, H., Korbin, S., & Chollete, V. (2007). What do women in the US know about human papillomavirus and cervical cancer? *Cancer Epidemiological Biomarkers*, 16(2), 288–294.
- Toth, J. C. (2003). Measuring stress after acute myocardial infarction: The Stress of Discharge Assessment Tool, Version Two (SDAT2). In O. L. Strickland & C. Dilorio (Eds.), *Measurement of nursing outcomes* (2nd ed.), (Vol. 2, pp. 99–109). New York: Springer Publishing.
- Turris, S. A. (2009). Women's decisions to seek treatment for the symptoms of potential cardiac illness. *Journal of Nursing Scholarship*, 41(1), 5–12.
- Vedich, A. J., & Shapiro, F. (1955). A comparison of participant observation and survey data. *American Sociological Review*, 20, 28–33.
- Vernon, P. E., & Millican, G. D. (1954). A further study of the reliability of English essays. *British Journal of Statistical Psychology*, 7, 65–74.
- Waltz, C. F. (1988). *Educational outcomes: Assessment of quality—A prototype for student outcome measurement in nursing programs*. New York: National League for Nursing Publications.
- Webb, E. E., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.
- Weber, J., Kopelman, S., & Messick, D. (2004). A conceptual review of decision making in social dilemmas: Applying a logic of appropriateness. *Personality and Social Psychology Review*, 8(3), 281–307.
- Williams, K. N. (2006). Improving outcomes of nursing home interactions. *Research in Nursing and Health*, 29(2), 121–133.