Intervention Research and Evidence-Based Quality Improvement
Designing, Conducting, Analyzing, and Funding

Bernadette Mazurek Melnyk, PhD, RN, APNP-BC, FAANP, FNAP, FAAN
Dianne Morrison-Beedy, PhD, RN, WHNP, FAANP, FNAP, FAAN

EDITORS

First Edition received 2012 First Place AJN Book of the Year Award in Nursing Research!

“This is a resource for success and should be a part of any researcher’s library.”
—Doody’s Medical Reviews (Praise for the First Edition)

Written for researchers, clinicians, and doctoral students, the newly revised edition of this comprehensive reference continues to deliver the essentials of intervention research with added content on evidence-based quality improvement, essential for improving healthcare quality, safety, and population health outcomes. Although typically it takes years for research-based interventions to make their way to real-world clinical settings, this prolonged time for translation frustrates researchers and their interprofessional teams. This second edition now delves even deeper into key strategies for rapidly moving research-based interventions into real-world settings in the form of evidence-based quality improvement as well as the challenges of working in an increasingly diverse professional research environment.

Intervention Research and Evidence-Based Quality Improvement, Second Edition, begins at the pilot study phase for intervention research and highlights every step of the way through to full-scale randomized controlled trials. Written in a user-friendly format, content covers designing, conducting, analyzing, and funding intervention studies that improve healthcare quality and people’s health outcomes. Chapters cover writing grant applications and show examples of actual applications that have been funded by the National Institutes of Health and other organizations. These real-life samples are available online, alongside additional progress reports and final reports. Real-world examples of evidence-based quality improvement projects that have improved outcomes also are highlighted in this second edition.

New to the Second Edition:
- Describes evidence-based quality improvement and specific steps in conducting EBQI projects, which are essential for enhancing healthcare quality, safety, and costs along with enhancing population health outcomes
- Emphasizes the importance of interprofessional teams
- Focusing on using research-based interventions in real-world settings
- Six new chapters
  - Generating Versus Using Evidence to Guide Best Practice
  - Setting the Stage for Intervention Research and Evidence-Based Quality Improvement
  - Evidence-Based Quality Improvement
  - Translational Research: Why and How
  - Factors Influencing Successful Uptake of Evidence-Based Interventions in Clinical Practice
  - Using Social Media to Enhance Uptake of Research-Based Interventions into Real World Clinical Settings

Key Features:
- Provides a practical, comprehensive resource for designing, conducting, analyzing, and funding intervention studies
- Outlines the specific steps in designing, conducting, and evaluating outcomes of evidence-based quality improvement projects
- Includes examples of funding research grants, progress reports, and final reports
- Serves as a core text for students in doctoral nursing and other health sciences programs

Second Edition

Intervention Research and Evidence-Based Quality Improvement
Designing, Conducting, Analyzing, and Funding

Melnyk  •  Morrison-Beedy

SPRINGER PUBLISHING

11 W. 42nd Street
New York, NY 10036-8002
www.springerpub.com
Intervention Research and Evidence-Based Quality Improvement
Bernadette Mazurek Melnyk, PhD, RN, APRN-CNP, FAANP, FNAP, FAAN, is the vice president for health promotion, university chief wellness officer, professor and dean of the College of Nursing at The Ohio State University (OSU), and professor of pediatrics and psychiatry at OSU’s College of Medicine. She is a nationally/ internationally recognized expert and researcher in intervention studies, evidence-based practice (EBP), child and adolescent mental health, and health and wellness, and is a frequent keynote speaker at national/international conferences on these topics. Dr. Melnyk has led the EBP movement in nursing and healthcare across the United States for the past two decades, and has consulted with hundreds of healthcare systems and academic institutions throughout the nation and globe on how to improve healthcare quality and population health outcomes through EBP. Her record as principal investigator includes more than $36 million of sponsored funding from federal agencies and foundations. She is editor of five books and has more than 345 publications, including two American Journal of Nursing Book of the Year awards. She is an elected fellow of the National Academy of Medicine, the American Academy of Nursing, the National Academies of Practice, and the American Association of Nurse Practitioners. She served a 4-year term on the United States Preventive Services Task Force and the National Institutes of Health (NIH) National Advisory Council for Nursing Research. Dr. Melnyk is currently a member of the National Quality Forum’s Behavioral Health Steering Committee and the National Academy of Medicine’s Action Collaborative on Clinician Wellbeing. She also serves as editor of the journal Worldviews on Evidence-based Nursing. Dr. Melnyk has received numerous national and international awards, including the inaugural National Institutes of Health/National Institute of Nursing Research Director’s Distinguished Lectureship Award, the Lifetime Achievement Award by the National Organization of Nurse Practitioner Faculty, the Loretta Ford Award from the National Association of Pediatric Nurse Practitioners, and the Nancy Sharpe Cutting-Edge Award by the American Association of Nurse Practitioners. She also has been twice named an Edge Runner by the American Academy of Nursing and was inducted into Sigma Theta Tau International’s Research Hall of Fame. Dr. Melnyk is the founder and executive director of the Helene Fuld Health Trust National Institute for Evidence-based Practice in Nursing & Healthcare, which is housed at the OSU College of Nursing. Dr. Melnyk founded the National Interprofessional Education and Practice Collaborative to advance the Department of Health and Human Services’ Million Hearts® initiative to prevent 1 million heart attacks and strokes by 2017, which now has more than 150 participating academic institutions and organizations throughout the United States. She also created and chaired the first three National Summits on Building Healthy Academic Communities and founded the National Consortium for Building Healthy Academic Communities, a collaborative organization to improve population health in the nation’s institutions of higher learning, for which she served as its first president.

Dianne Morrison-Beedy, PhD, RN, WHNP, FNAP, FAANP, FAAN, currently serves as chief Talent and Global Strategy officer and the Centennial Endowed Professor of Nursing at The Ohio State University. Prior to that, she was the dean of the University of South Florida (USF) College of Nursing, senior associate vice president of USF Health, and a professor of nursing, global health, and public health at USF. Dr. Morrison-Beedy has focused her research on HIV/AIDS risk reduction, especially for vulnerable adolescent girls. She has received more than $11 million in HIV prevention research funding and was inducted into the Sigma Theta Tau International Nursing Research Hall of Fame for developing an evidence-based sexual risk reduction intervention, the Health Improvement Project for Teens (HIPTeens), recognized by the Centers for Disease Control and the U.S. Department of Health and Human Services for its strong evidence for HIV/STI and pregnancy prevention. Internationally recognized as a scholar, administrator, and educator, she recently received two Fulbright awards—an administrator award (France) and a research scholar award (Scotland)—as well as an International Council of Nurses Global Leadership fellowship (Switzerland). Dr. Morrison-Beedy’s interdisciplinary contributions also encompass serving as scientific reviewer for multiple HIV-related study sections and special emphasis panels at the NIH. In recognition of her contributions, she has been awarded the Florida Nursing Association Award for Research and the Association of Women’s Health, Obstetric and Neonatal Nurses Award for Excellence in Research, the Association of Nurses in AIDS Care (ANAC) Research Recognition Award, Excellence in HIV Prevention Award, and the New York State Distinguished Nurse Researcher Award. She is an elected fellow in the American Academy of Nursing, the American Academy of Nurse Practitioners, and the National Academies of Practice. Dr. Morrison-Beedy served as faculty for Sigma Theta Tau International Chiron Nursing Leadership Program and Johnson & Johnson Maternal and Child Health Leadership Institute. She is a reviewer for numerous scientific journals and has more than 250 published articles and abstracts. Dr. Morrison-Beedy received her bachelor’s degree from Niagara University, her master’s degree with a specialization in maternal and women’s health nurse practitioner from the State University of New York at Buffalo, and her doctoral degree at the University of Rochester.
Contents

Contributors ix
Foreword xiii
Preface xv
Acknowledgments xvii

Share Intervention Research and Evidence-Based Quality Improvement: Designing, Conducting, Analyzing, and Funding, Second Edition

PART I: SETTING THE STAGE

1. Generating Evidence Through Intervention Research Versus Using Evidence in Evidence-Based Practice/Quality Improvement 3
Bernadette Mazurek Melnyk and Dianne Morrison-Beedy

2. Setting the Stage for Intervention Research and Evidence-Based Quality Improvement: The “So What,” “What Exists,” and “What’s Next” Factors 11
Bernadette Mazurek Melnyk and Dianne Morrison-Beedy

3. Designing, Conducting, and Evaluating Outcomes of Evidence-Based Quality Improvement Projects 21
Bernadette Mazurek Melnyk

PART II: DESIGNING INTERVENTION STUDIES

4. Using Theory to Guide Intervention Research 55
Julie Fleury and Souraya Sidani

5. Nuts and Bolts of Designing Intervention Studies 79
Bernadette Mazurek Melnyk, Dianne Morrison-Beedy, and Shirley M. Moore

6. Designing Interventions That Are Sensitive to Culture, Race/Ethnicity, and Gender 105
Usha Menon

7. Ethical Considerations in Designing Intervention Studies 115
Dónal P. O’Mathúna

8. Minimizing Threats to Internal Validity 129
Jacqueline Dunbar-Jacob

©Springer Publishing Company
### Part III: Implementing Intervention Studies

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Developing the Budget for Intervention Studies</td>
<td>185</td>
</tr>
<tr>
<td>13</td>
<td>Navigating the IRB for Investigators</td>
<td>203</td>
</tr>
<tr>
<td>14</td>
<td>Participant Recruitment and Retention</td>
<td>225</td>
</tr>
<tr>
<td>15</td>
<td>Maintaining Fidelity of an Intervention</td>
<td>245</td>
</tr>
<tr>
<td>16</td>
<td>Study Implementation: An Example Using the Madres Para la Salud (Mothers for Health) Study</td>
<td>261</td>
</tr>
<tr>
<td>17</td>
<td>Considerations in Conducting Multisite Trials</td>
<td>275</td>
</tr>
<tr>
<td>18</td>
<td>Considerations in Conducting Interventions in Specialized Settings</td>
<td>281</td>
</tr>
<tr>
<td>18.1</td>
<td>Conducting Interventions in Global Settings</td>
<td>282</td>
</tr>
<tr>
<td>18.2</td>
<td>Conducting Interventions in Long-Term Care Settings</td>
<td>292</td>
</tr>
<tr>
<td>18.3</td>
<td>Conducting Interventions in Public Health Settings</td>
<td>298</td>
</tr>
<tr>
<td>18.4</td>
<td>Conducting Interventions in School Settings</td>
<td>306</td>
</tr>
<tr>
<td>18.5</td>
<td>Conducting Interventions in Palliative Care Settings</td>
<td>310</td>
</tr>
<tr>
<td>18.6</td>
<td>Conducting Interventions in Acute Care Settings</td>
<td>315</td>
</tr>
<tr>
<td>19</td>
<td>Data Management</td>
<td>327</td>
</tr>
</tbody>
</table>

© Springer Publishing Company
PART IV: ANALYZING INTERVENTION STUDIES

20. Analyzing Intervention Studies 349
   Laura A. Szalacha

21. Cost-Effectiveness Analyses for Intervention Studies 363
   Kimberly Arcoleo and Kevin Frick

22. Explaining Intervention Effects 377
   Kimberly Arcoleo

PART V: WRITING SUCCESSFUL GRANTS AND PROGRESS REPORTS

23. Writing Grants That Fund! 395
   Bernadette Mazurek Melnyk and Dianne Morrison-Beedy

24. Submitting a Research Grant Application to the National Institutes of Health:
   Navigating the Application and Peer Review System 417
   Mindy B. Tinkle and Ann Marie McCarthy

25. Writing Progress and Final Reports 433
   Margaret Roudebush and Shirley M. Moore

PART VI: MOVING INTERVENTION FINDINGS INTO THE REAL WORLD

26. Disseminating Findings of Intervention Studies 459
   Marion E. Broome

27. Social Media Use in Intervention Research and Dissemination 483
   Linsey N. Grove and Dianne Morrison-Beedy

28. Speeding the Translation of Evidence-Based Interventions Into Real-World Practice Settings 493
   Bernadette Mazurek Melnyk

29. Factors Influencing Successful Uptake of Evidence-Based Interventions in Clinical Practice 503
   Sharon Tucker

Glossary 519
Index 533
Contributors

Barbara E. Ainsworth, PhD, MPH, FACSM, FNAK  Regents’ Professor, School of Nutrition and Health Promotion, Arizona State University, Queen Creek, Arizona

Kimberly Arcoleo, PhD, MPH  Associate Dean for Research, University of Rochester School of Nursing, Rochester, New York

Carol M. Baldwin, PhD, RN, CHTP, CT, NCC-BC, AHN-BC, FAAN  Professor Emeritus; Southwest Borderlands Scholar, Deputy Director, PAHO/WHO Collaborating Centre to Advance the Policy on Research for Health, College of Nursing & Health Innovation, Arizona State University, Phoenix, Arizona

Jacqueline Dunbar-Jacob, PhD, RN, FAAN  Dean and Distinguished Service Professor of Nursing, University of Pittsburgh School of Nursing, Pittsburgh, Pennsylvania

Julie Fleury, PhD, FAAN, FAHA  Hanner Memorial Endowed Professor, Guy Hanner Professor of Nursing, College of Nursing and Health Innovation, Arizona State University, Phoenix, Arizona

Kevin D. Frick, PhD  Professor Johns Hopkins Carey Business School, Baltimore, Maryland

Bonnie Gance-Cleveland, PhD, PNP, RNC, FAAN  Loretta C. Ford Professor, University of Colorado, Morrison, Colorado

Sandra Glover Gagnon, PhD  Associate Professor, Department of Psychology, Appalachian State University, Boone, North Carolina

Linsey Grove, MPH, CPH, CHES  Chief Creative Officer and Owner, Eunoia Media Lab, LLC, St. Petersburg, Florida

Mary Beth Happ, PhD, RN, FAAN  Associate Dean of Research and Innovation, Distinguished Professor of Critical Care Research, The Ohio State University College of Nursing, Columbus, Ohio

Colleen Keller, PhD, RN, FNP, FAAN  Regents’ Professor and Foundation Professor in Women’s Health, Arizona State University College of Nursing and Health Innovation, Phoenix, Arizona
Kevin E. Kip, PhD  Interim Chair, Distinguished USF Health Professor, University of South Florida, Tampa College of Public Health, Florida

Catherine G. Ling, PhD, FNP-BC, FAANP  Associate Professor, Associate Dean for Graduate Clinical Programs, USF Health College of Nursing, Tampa, Florida

Ann Marie McCarthy, PhD, RN, FNASN, FAAN  Professor, Associate Dean for Research and Scholarship, College of Nursing, University of Iowa, Iowa City, Iowa

Sandee G. McClowry, PhD, RN, FAAN  Professor, New York University, New York

Usha Menon, PhD, RN, FAAN  Professor and Associate Dean of Research & Global Advances, University of Arizona College of Nursing, Tucson, Arizona

Shirley M. Moore, PhD, RN, FAAN  The Edward J. and Louise Mellen Professor of Nursing, Distinguished University Professor, Associate Dean for Research, Frances Payne Bolton School of Nursing Case Western Reserve University, Cleveland, Ohio

Cindy L. Munro, PhD, APRN-BC  Associate Dean of Research, University of South Florida College of Nursing, Apollo Beach, Florida

LaRon E. Nelson, PhD, RN, FNP, FNAP, FAAN  Scientist & OHTN Applied HIV Research Chair, St. Michael’s Hospital, Li Ka Shing Knowledge Institute, Centre for Urban Health Solutions, Toronto, Ontario, Canada

Debra Parker Oliver, PhD, MSW  Paul Revare, MD, Family Professor of Family Medicine, School of Medicine, University of Missouri Health, Columbia, Missouri

Dónal P. O’Mathúna, PhD, MA, BSc (Pharm)  Associate Professor, School of Nursing & Human Sciences, Dublin City University, Ireland, and College of Nursing, The Ohio State University, Columbus, Ohio

Steven E. Pease, BS, MAS  Assistant Dean, Research and Finance Administration, University of Maryland School of Nursing, Baltimore, Maryland

Marlys R. Peck, PhD, MSW  Associate Professor, Central Missouri State University, Sedalia, Missouri

Alicia G. Rossiter, DNP, FNP, PCNP-BC, FAANP  Assistant Professor Director, Veteran to BSN Program, College of Nursing Military Liaison, University of South Florida College of Nursing, Tampa, Florida

Margaret Roudebush, MNO  Assistant Dean for Research Administration, Center for Research and Scholarship, Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, Ohio

Barbara A. Smith, PhD, RN, FAAN  Professor and Interim Vice Dean of Academic Affairs, University of South Florida, Tampa, Florida, and Professor, Florida State University, Tallahassee, Florida
Laura A. Szalacha, EdD  Director of Research Methods and Statistics, Professor, University of Arizona College of Nursing, Tucson, Arizona

Mindy B. Tinkle, PhD, RN, WHNP-BC, FAAN  Associate Professor, University of New Mexico College of Nursing, Albuquerque, New Mexico

Sharon Tucker, PhD, RN, FAAN  Grayce Sills Endowed Professor in Psychiatric-Mental Health Nursing and Director of the Implementation Science and Systematic Review Core, The Helene Fuld Health Trust National Institute for Evidence-Based Practice in Nursing and Healthcare College of Nursing, The Ohio State University, Columbus, Ohio

Peter A. Vanable, PhD  Associate Provost for Graduate Studies, Dean of the Graduate School, Professor of Psychology, Syracuse, New York
Foreword

As we near the end of the second decade of the millennium, a focus on rigorous and relevant intervention research and evidence-based practice continues to be an essential component of our nation’s efforts to achieve a high-performing learning health system. Despite progress in reducing the number of uninsured and significant public and private sectors investments in improvement activities, we are still failing to provide consistent, high-quality, safe care to everyone. In fact, the latest National Healthcare Quality and Disparities Report from the Agency for Healthcare Research and Quality underscores the need: Progress is seen for only 52% of quality measures for which trend data are available and disparities in care are only improving for 16% of outcome measures.

We must do better, and this comprehensive updated volume from Melnyk and Morrison-Beedy will be a tremendous resource for researchers and clinicians alike. Its detailed, step-by-step approach to designing, implementing, analyzing, and scaling interventions into real-world settings will assist novice and skilled researchers alike. The inclusion of comparative effectiveness research is notable as health systems, clinicians, and patients want and need to make choices between available alternatives—understanding which intervention “works” for which patients in which settings. Of particular value is the attention to interventions that are sensitive to culture, race/ethnicity, and gender as well as the inclusion of considerations of diverse care settings. At the same time, as healthcare costs continue to rise and challenge payers, providers, and patients, we must heed the call from Melnyk and colleagues to focus on the “so what” questions. We should target our interventions to achieve the most impact. Evidence-based quality improvement requires dedicated resources from health systems and understanding the costs and cost-effectiveness—a critical aspect too often ignored by experts—is addressed in this volume.

The next decade of intervention studies and evidence-based quality improvement must not build just on what we know, but capitalize on the many opportunities for disruptive innovation and leveraging the increasing volume, variety, and velocity of data types. These forces, together with the trend toward multiside studies and team science, present significant opportunities for intervention research and evidence-based quality improvement. There also is a growing community of researchers and health system leaders impatient with the status quo and eager to partner to achieve better care for their patients. This growth is evident each year at the Conference on the Science of Dissemination and Implementation in Health as well as in the increasing number of healthcare delivery science centers being established in health systems and academic health centers. Every one of them should have a copy of this volume!

Lisa Simpson, MB, BCh, MPH, FAAP
President and CEO
AcademyHealth
Imagine you hear about a wonderful product that has just been developed that could enhance your quality of life; it is far better than what is available to you currently. The word on the street is that people really need this product—it could change their lives for the better. You, your friends, and your family are really interested in getting access to this product. There is only one issue: it will not be available to you or others for 10, 15, or maybe 20 years. Disappointed? Upset? Frustrated? Unfortunately, this is often the typical scenario with evidence-based interventions that have been rigorously tested and found to be effective through research. It typically takes years, even decades, to translate research-based interventions into real-world clinical settings to improve healthcare quality and people’s health outcomes. This long research–practice time gap does not sit well with either of us as we are truly passionate about making a positive difference in the health and well-being of people, families, and communities through intervention, development, testing, and translation. As friends and colleagues for almost three decades, we had the opportunity to have multiple discussions about how best to conduct meaningful and rigorous research that would target these unmet needs. After our doctoral programs, both of us pursued our dreams of making a difference through rigorous programs of intervention research. We went through the typical “character builders” of every step involved in intervention work, beginning with pilot studies that were eventually funded by the National Institutes of Health (NIH) as full-scale randomized controlled trials.

The first edition of our book, Intervention Research: Designing, Conducting, Analyzing, and Funding, included many of the lessons that we learned and tricks of the trade that we compiled into one user-friendly reference where others could learn from those who have successfully traveled down the intervention pathway. We were grateful that the first edition met your expectations and was also recognized by the American Journal of Nursing as the Research Book of the Year in 2012. In this second edition, we knew we had to extend the reach and impact of the book by addressing the ongoing dilemma of the “prolonged time to translation” of intervention research to clinical practice. Therefore, additional content on how to move research-based interventions more quickly into real-world settings has been added. Supplementary appendices are available at www.springerpub.com/melnyk. Qualified instructors may obtain access to an ancillary instructor’s manual and instructor’s PowerPoints by emailing textbook@springerpub.com.

With the current confusion that exists in academia and healthcare regarding the preparation and role of individuals with clinical doctorates (e.g., the Doctor of Nursing Practice [DNP]), we begin this new edition with tackling the difference between “generating evidence” through intervention research versus “using evidence” through evidence-based quality improvement. Therefore, whether your passion lies in generating
versus using evidence to improve real-world outcomes, we hope that this book will assist you in realizing your dreams through intervention work and evidence-based quality improvement.

Abraham Lincoln said that the best way to predict the future is to create it; so keep on dreaming, discovering, and delivering to transform healthcare and improve lives!

Warm regards,

Bernadette “Bern” Mazurek Melnyk and  
Dianne Morrison-Beedy
Acknowledgments

It usually takes an awesome team to accomplish a major initiative, such as the production of a book. Thanks and recognition go to my wonderful friend, colleague, and coeditor, Dr. Dianne Morrison-Beedy, who embarked on this dream and journey with me, weathering the “character-builders” along the course of the way. I also want to thank and acknowledge each of our terrific contributors, without whom this book would not have been possible. In addition, I want to recognize my supportive husband, John, and my three wonderful daughters, Kaylin, Angela, and Megan, who sacrificed time with me over the years so that I could pursue my dreams. Kathy York also deserves special recognition because she has been the “wind beneath my wings” for more than a decade, as well as my colleagues, study teams, students, and research participants from whom I have learned a lot. Finally, I credit two incredibly special people for my success—my dad Joe Mazurek, who gave me a spirit of never-ending determination, and my sister Chris, who was my “lifeline” during many of my life’s challenges and who always encouraged me to “just get out there and do it!”

Bernadette Mazurek Melnyk

For this edition of our book, I am thinking a bit broader and deeper about acknowledgements. To my friend, colleague, idea partner, and coeditor Bernadette—30 years later and we are still such a good team. It is always a pleasure to bring our many dreams to reality. Also, at the top of the list are my dear husband Michael, son Mason, and daughter Megan. They are my life and my inspiration for trying to be part of the “change in the world” I am hoping to see in nursing and healthcare. I also thank my parents who always told me that I could do great things in this world that may require much effort (witness this book), but my dreams were possible. Now to you, our readers—my acknowledgement is to each of you. You are reading this book to gain a better understanding of how your creative ideas combine with rigorous science to produce an impact on real people. I applaud you on your efforts. Keep going in the face of adversity, grant rejections, manuscript revisions, or idea bashing. You will come out successful on the other end! What I have learned over the past year, however, is this: take care of yourself as you pursue your dreams; no one can really do that for you. All the essentials of health we try to convey and encourage our patients and families to undertake are just as important for us as individuals. So if you have not made an appointment for a checkup in a long time, make it. If you are thinking about having a snack, take the time to make yourself a delicious and healthy meal. If you need a break from reading this book, get up and take a walk. If you need to catch up on sleep, it is okay to put work away and go to bed early. Take care of yourself. This way you will still be around to contribute to research that transforms health and healthcare (and read our next edition!). Stay healthy and happy—you need to do that for you.

Dianne Morrison-Beedy

©Springer Publishing Company
Share

Intervention Research and Evidence-Based Quality Improvement: Designing, Conducting, Analyzing, and Funding, Second Edition
Using Theory to Guide Intervention Research

Julie Fleury & Souraya Sidani

The important thing is to never stop questioning. — Albert Einstein

Theory has an integral role in the design of interventions and studies aimed at evaluating the effects of interventions consisting of single or multiple components. Theory provides an understanding of the problem that the intervention targets, the nature of the intervention and/or its components, and the mechanisms underlying the anticipated improvement in outcomes. It also guides sample selection, operationalization of the intervention, and the specification of outcomes. Results of theory-based intervention evaluation indicate who benefits from the intervention, delivered in what context, and how the intervention and/or its components produce(s) changes in the outcomes. This knowledge is useful for translating research findings to practice and in making relevant decisions regarding the application of the intervention and/or its components in the practice setting. In this chapter, the role of theory in intervention studies is reviewed. Strategies for developing interventions are briefly described. The contribution of theory to the design of interventions and pertinent evaluation studies is clarified and illustrated with examples. Finally, sources for generating theory are identified and illustrated with examples from the literature.

ROLE OF THEORY IN INTERVENTION STUDIES

Leaders in social, behavioral, and health sciences have emphasized the importance of theory as a guide for intervention design and evaluation. At the simplest level, a theory is an explanation of why a phenomenon occurs the way it does. Theory reflects a body of knowledge that organizes, describes, predicts, and explains a phenomenon. In intervention
research, theory or theoretical frameworks allow us to better represent the complexity of the situation under study. Frameworks clarify the problem targeted by the intervention, providing information on the determinants or factors causing the problem, the population experiencing the problem, the specific intervention strategies likely to address the determinants and/or to alleviate or improve management of the problem, the important steps or links of the transformation process the intervention brings about, and the specific outcomes indicating the effectiveness of intervention. Thus, theoretical frameworks guide the development of interventions that are more likely to permit strong causal inference and interpretable results (Sidani, 2015). This information is essential for developing and expanding the knowledge base for clinical practice and assisting clinicians in selecting and prescribing the most appropriate interventions.

Theoretical frameworks foster a systematic approach to intervention development and implementation that allow us to move beyond a simplistic, outcomes-focused approach to examining the central processes underlying intervention effects (Prestwich et al., 2014). In many instances, intervention effects are observed, but conclusions underlying the mechanisms for change remain isolated, without comprehensive understanding of the theory underlying intervention development, implementation, and evaluation. Theory-based interventions are central to addressing, in a sound and responsive manner, needs in the promotion of health, prevention, and management of illness as well as eliminating disparities in health and healthcare among vulnerable and underserved populations.

The use of theory as a basis for intervention design stems from an internally consistent group of relational statements specifying how the problem of interest comes about, and how that problem may be prevented, managed, or resolved. Theoretical frameworks guiding interventions are directly testable, including an integrated set of concepts, existence statements, and relational statements that link the problem to be addressed with relevant intervention strategies and the expected series of changes leading to the achievement of intended outcomes. This approach enhances sensitivity in intervention design by decreasing trial and error in design and implementation. In addition, theoretical frameworks direct the specification of various aspects of an intervention evaluation study, which contributes to enhanced validity of conclusions. Use of theoretical frameworks enables the investigator to identify what constitutes the intervention and what does not, what components or elements of the intervention are crucial, and where variation is possible, reducing the possibility of alternative rival hypotheses and strengthening internal validity (Sidani, 2015). A theory-based approach enhances construct validity through clear operationalization of the intervention, which enhances fidelity of implementation, and through specification of the intervention goals from which the outcomes are derived. Accordingly, the selected outcomes are sensitive to the intervention and the likelihood of detecting significant hypothesized effects is improved.

**Theoretical frameworks** are effective guides for research, practice, and the development of interventions that improve the health of the public. As a guide to intervention research, theoretical specificity allows for (a) accurate conceptualization of the problem targeted by the intervention, as well as specification of the population responsive to the intervention, (b) specification of the critical inputs that operationalize the intervention, (c) delineation of contextual factors that influence the implementation and outcomes of the intervention, including characteristics of the target population, interventionists, and setting or environment that moderate the effects of the intervention on outcomes, and (d) understanding of the mechanisms or processes of change yielding the expected or desired outcomes, including identification of specific mediators of change that can determine why an intervention was or was not successful.

Despite the acknowledged strengths of theory-based intervention research, the explication of theory in intervention development and evaluation and, in particular, the theoretically predicted mechanisms of change have been limited. Conn, Cooper, Ruppar, and Russell (2008)
reviewed the published reports of 141 studies that evaluated nursing interventions. They found that, in general, the interventions under evaluation were incompletely described; 27% reported enough detailed information to replicate the intervention in another study or translate the intervention into practice. These authors indicate that although about half (52%) of the studies referenced a theoretical framework on which the intervention under evaluation was based, there was minimal explanation of the linkages between the concepts in the framework and the content of the intervention. In response to these concerns, Michie and Prestwich (2010) developed an objective approach for assessing the extent to which behavioral interventions are theory-based, as well as the degree to which intervention development and evaluation are informed by theory. Similarly, there are a very limited number of studies that clearly articulate the mechanism explaining the effects of the intervention on the selected outcomes. Hoffmann et al. (2014) introduced the Template for Intervention Description and Replication (TIDieR), an approach for specifying dimensions of interventions including characterizing the delivery and the content domain of activities that comprise the intervention.

Understanding the problem requiring management, the specific aspects of the problem addressed by the intervention, the components or elements characterizing the intervention, and the mechanism or processes underlying its effects are critical for implementing the intervention and attaining the beneficial outcomes in the context of practice. Such an understanding is generated through the application of a careful and systematic process for developing interventions.

**STRATEGIES FOR INTERVENTION DEVELOPMENT**

There is limited consensus on the processes central to the development of theory-based interventions. Recent frameworks have been proposed, including the Medical Research Council (MRC) framework (Craig et al., 2013), intervention mapping (IM; Kok et al., 2016), behavioral change techniques (BCT; Michie et al., 2016), and the realist approach to evaluation (Pawson & Manzano-Santaella, 2012). In addition, the multiphase optimization strategy trial (MOST) has been suggested as a strategy to build multicomponent interventions and optimize their effectiveness (Collins, Dziak, Kugler, & Trail, 2014). Although useful in providing investigators with a structured process to map intervention mechanisms to relevant aspects of the problem requiring management and to identify active intervention components that best contribute to positive outcomes, these frameworks and strategies do not explicitly explain the role of theory in intervention design. The problem-solving approach represents a useful alternative that emphasizes the clear delineation of the links between the problem and the elements or components of the intervention carefully selected to address the problem; these links also inform the specification of the mechanisms underlying the intervention effects on outcomes.

**Medical Research Council Framework**

The MRC framework (Craig et al., 2013) addresses approaches for developing and evaluating complex interventions using a stepwise approach starting with identifying existing evidence in support of the intervention, identifying and developing theory, and progressing to modeling, feasibility, evaluation, and implementation phases. In the MRC framework, context is central to intervention design and evaluation, including socioeconomic background, the health service systems, the characteristics of the population, the prevalence or severity of the condition studied, and how these factors change over time. Identifying existing evidence in support of the intervention clarifies what is known about similar interventions and evaluation methods. Identifying and developing theory reflects an understanding of the processes of change, drawing on empirical data and supported by preliminary research. Modeling
involves hypothesizing and testing both what to target (e.g., behavioral determinants of the problem) and how to do this (e.g., techniques to change these determinants). Modeling of the intervention both depends on, and informs, understanding of the underlying problem. Feasibility approaches are recommended to explore key uncertainties regarding intervention acceptability, compliance, delivery of the intervention, participant recruitment and retention, and proposed effect size. Evaluation includes selection of methodological choices, including primary and secondary outcome variables. Process evaluation is used to address fidelity of intervention implementation, clarify causal mechanisms, and identify contextual factors associated with variation in outcomes. Additional key tasks include identifying barriers to application of the intervention and considering the best achievable combination of intervention components and intensity. Implementation includes dissemination of results as widely as possible, with further research to assist and monitor the process of implementation.

**Intervention Mapping**
Kok et al. (2016) describe a process for intervention development, implementation, and evaluation of behavioral change interventions. This process integrates theory and evidence in (a) assessment of needs or problem analysis identifying what needs to be changed and for whom; (b) development of matrices of program objectives reflecting behaviors and behavioral determinants; (c) selection of theory-based intervention methods and practical strategies to change behavioral risk factors for health problems; (d) production of program components; (e) anticipation of program adoption, implementation, and sustainability; and (f) anticipation of process and effect evaluation. IM outlines a process from the recognition of a need or problem to the identification of a solution. Products such as program matrices, planning documents, and program materials are developed throughout each step of the process. The role of theory is most relevant during the phase of selecting program strategies relevant to each program objective where theory may guide the choice or the development of strategies to accomplish program objectives. Accordingly, interventions based on the IM approach do not have an exclusive focus on one theory, but rather they reflect a combination of the most promising theoretically derived intervention strategies for a given problem.

**Behavioral Change Techniques**
This approach is designed to clarify the link between intervention BCT and mechanisms of action (Michie et al., 2016). The BCT framework is designed to characterize interventions as well as provide a system for matching intervention strategies to a behavioral target, a target population, and the context in which the intervention will be delivered. The BCT approach has included a causal analysis of behavior and conditions specific to individuals in their social and physical environment, and the changes needed for a specified behavioral target to be achieved. Interventions are designed and selected according to an analysis of the nature of the behavior, the mechanisms that need to be changed to bring about behavioral change, and the interventions needed to change these mechanisms. Michie and Prestwich (2010) present a theory-coding scheme for assessing the extent to which theory has been applied to developing and evaluating interventions targeting behavioral change. The coding scheme includes six categories related to (a) reference to theory in intervention design, (b) targeting relevant theoretical constructs in intervention strategies, (c) using theory to select participants or tailor interventions, (d) measurement of relevant theoretical constructs, (e) testing mediation effects, and (f) refining theory. Understanding how theory has been applied and tested in a given body of literature provides empirical support for theory selection. Michie et al. (2015) introduced the BCT Taxonomy, a formalized approach to characterizing BCT. BCT are operationally defined to facilitate a common dialogue and description. Current research is focused on developing matrices of hypothesized links between behavioral change techniques, specific mechanisms of action, and behavioral theory (Michie et al., 2016).
Realist Evaluation Approach

The realist approach is a theory-driven strategy for evaluating single- or multiple-component interventions. It is concerned with determining not only the net effects of an intervention on outcomes, but also with generating an understanding of what intervention or components work, for whom, in what circumstances, in what respects, over which duration, and how or why (Pawson & Manzano-Santaella, 2012; Salter & Kothari, 2014).

The realist approach to evaluation assumes there is an underlying theory behind the working of an intervention. The middle-range theory explains variations in outcomes observed following the interventions relative to mechanisms and contexts; it is configured as Context–Mechanism–Outcome (CMO). Mechanisms represent how the intervention and its components are received, interpreted, and acted upon by patients to produce the outcomes (Dalkin, Greenhalgh, Jones, Cunningham, & Lhussier, 2015). Contexts are the environmental (physical and sociopolitical) and personal (physical, psychological, and sociocultural) factors that affect the implementation of the intervention, the patients’ reactions and responses to the intervention, and hence, the changes in intended outcomes (Bonell, Fletcher, Morton, Lorenc, & Moore, 2012). The outcomes reflect improvement in patients’ health problems or behaviors.

The application of the realist evaluation approach requires the development of the just mentioned middle-range theories before or as part of the evaluation study. This can be accomplished by reviewing and synthesizing relevant theoretical and empirical literature that informs a priori generation of the theory to guide the design and conduct of the study, and/or by incorporating qualitative methods (e.g., interviews with the persons who developed, implemented, and received the intervention) to explore the contextual factors and the mechanisms underlying the effectiveness of the intervention (Pawson & Manzano-Santaella, 2012).

Multiphase Optimization Strategy

The multiphase optimization strategy trial (MOST) is used to optimize the effectiveness of the intervention. This is done by first dismantling the intervention into its components; these include the specific and nonspecific elements that contribute to the outcomes. Second, factorial or fractional factorial designs are carried out to identify (a) the components that are active and lead to best outcomes, (b) the components that are inactive and have minimal or detrimental effects on outcomes, and (c) the extent to which one component impacts the effects of others. The results will help to revise the intervention to optimize its effects, by omitting ineffective components (Collins et al., 2014).

These frameworks and evaluation strategies provide an important guide for specifying intervention strategies and intended outcomes and offer general templates for mapping behavioral change intervention strategies to behavioral determinants. Building on these frameworks, efforts are needed to clearly articulate the nature of the links that are created when mapping intervention strategies to behavioral determinants, clarifying the mechanisms through which intervention strategies achieve effects. Without understanding these links, it will be difficult to derive the testable hypotheses, derived from the intervention middle-range theories, needed to further intervention science. Detail on the nature of the problem of interest, the context of the problem, how to use theory to conceptualize the problem of interest, and how to design an intervention responsive to the problem of interest allows greater specificity in intervention development and testing. Understanding the causal links between the problem of interest, intervention strategies, mediating variables, and intended outcomes permits an intervention to be delivered in a responsive manner, rather than in a mechanical or stereotypic format. Rothman et al. (2016) and Pawson and Manzano-Santaella (2012) have called for theory testing and refinement using a programmatic approach to clarify the underlying causal processes of behavioral change and the conditions under which these processes occur. To achieve this goal, an approach to intervention design is needed which furthers
knowledge about the mechanisms underlying a theory-based intervention and the problem it is designed to address. The problem-solving approach to intervention development, which has guided our work with colleagues and students for over a decade, represents such an approach.

**Problem-Solving Approach**

The problem-solving approach fosters understanding the problem requiring remediation; specifying populations experiencing the problem; clarifying the intervention in terms of critical inputs needed to address the problem, including components, implementation issues and exogenous factors; and identifying the expected outcomes of the intervention. These elements are integrated to form a theory that is used to explain for whom, why, and how the intervention works in producing the outcomes, under what circumstances, and to guide the design of an intervention evaluation study. The next sections discuss what each element is about and how it contributes to the study design. Exhibit 4.1 is a list of questions to be addressed when clarifying each element of the theory.

**EXHIBIT 4.1 The Problem-Solving Approach to Intervention Design**

### UNDERSTANDING THE PROBLEM OF INTEREST
- What are the manifestations and characteristics of the problem?
- What intensity, severity, and duration cause it to be labeled as a problem?
- What are the relevant antecedents to consider?
- When does the problem exist?
- What are the causative, associative, and contributing factors?
- At what level is the problem amenable to change?
- How do you want the problem to change—what are the desired outcomes?
- What would the change look like?
- Which theory best fits with the identified problem?
- How does the selected theory help to explain the problem?
- How does the theory identify how the problem comes about?
- How does the theory identify how the problem worsens or improves?
- How does the theory address antecedents and/or consequences of the problem?
- How does clinical knowledge support use of the theory?
- Does the theory reflect the experiences of the population?

### SPECIFYING THE POPULATION
- Who has the problem of interest?
- Who is susceptible?
- Under what conditions is that population susceptible?
- How will you measure the existence of the problem?
- How does the problem change over time, place, events, or people?

### MEDIATING PROCESSES
- Are the mediators proposed consistent with the problem of interest?
- Do the mediators explain how change will occur?
- Do the mediators clarify underlying mechanisms for change?
- Are the mediators composed of variables that are changeable?
- Is there a logical link between intervention critical inputs and mediators?
- Can the critical inputs be expected to have a direct effect on mediators?
- In what ways will critical inputs lead to change in mediators?

(continued)
EXHIBIT 4.1 The Problem-Solving Approach to Intervention Design (continued)

INTERVENTION THEORY—PROGRAM

■ What theory addresses the change desired?
■ What is the empirical support for the theory in designing intervention?
■ In what ways are the focus and assumptions of the intervention theory consistent with the problem theory?
■ Are there clear mechanisms for the change outlined consistent with the problem?
■ Are the nature and content of the intervention critical inputs clearly specified?
■ How do the specification of intervention critical inputs characterize what is necessary, what is sufficient, and what is optimal to produce the expected effects?
■ In what ways do the intervention critical inputs reflect theoretical foundations?
■ Is there a logical link between critical inputs and mediators?
■ Can the critical inputs be expected to have a direct effect on mediators?
■ In what ways will critical inputs lead to change in mediators?
■ Is the context for use of the theory clearly specified?
■ Is the theory relevant to the population of interest?
■ Is the theory gender and culturally relevant?
■ Is the theory developmentally appropriate?

STRENGTH AND DOSE

■ How are the strength and dosage of the intervention characterized?
■ What is the range within which the planned intervention is likely to show treatment effects?
■ What is the minimal level necessary to deliver the intervention at effective strength?

EXOGENOUS FACTORS

■ What are the individual factors that may affect treatment processes?
■ What are the contextual or environmental factors that may affect treatment processes?

IMPLEMENTATION ISSUES

■ What aspects of the treatment delivery system are relevant to consider?

TREATMENT FIDELITY STRATEGIES DESIGN

■ Is the treatment dose, including number, frequency, and length of contacts specified?
■ What strategies are in place to ensure a fixed dose in treatment delivery across participants and study conditions?
■ What strategies are in place to record deviations from protocol?
■ What mechanisms are in place to track the number, frequency, and duration of contacts?
■ Is there a comprehensive, scripted treatment manual developed with objectives and content for each session of treatment delivery?
■ What strategies are in place for externally monitoring sessions and providing feedback to interventionists?
■ What strategies are in place for interventionist self-monitoring of sessions?

INTERVENTIONIST TRAINING

■ How will training be standardized across interventionists?
■ How will skill acquisition in interventionists be measured?
■ How will change in interventionist skills be minimized?
INTERVENTION DELIVERY

- What mechanisms are in place to evaluate the extent to which interventions were delivered as intended regarding content and dose?
- How will contamination across interventions be minimized when delivered by the same interventionist?
- What mechanisms are in place to standardize intervention delivery?

INTERVENTION RECEIPT

- How will you verify that participants understand the information provided?
- How will you verify that participants can use the behavioral skills or evoke the subjective state they are trained to use?
- How will issues that interfere with receipt of the intervention be addressed?

ENACTMENT OF TREATMENT SKILLS

- What strategy can be applied to verify that participants use the cognitive, behavioral, and motivational skills and strategies provided in the intervention in the appropriate life situations?
- How will issues that interfere with enactment be addressed?

IDENTIFYING THE PRODUCT OF THE INTERVENTION

- What is the nature and form of the selected outcomes?
- What is the timing of treatment effects expected?
- What are the minimal magnitudes of effects thought to be clinically relevant?
- What is the maximal magnitude of effects thought to be likely?

CONTRIBUTION OF THE ELEMENTS OF INTERVENTION THEORY

Understanding the Problem

An important function of theory in intervention design is to guide the researcher in understanding the nature and characteristics of the problem targeted by the intervention. In general, there has been limited discussion in the literature around analyzing the nature of the problem of interest as the starting point for design of theory-based interventions. Therefore, one of the first steps in developing relevant theory-based interventions is the in-depth exploration of the problem in need of treatment.

Conceptualization of the problem specifies those “natural or social causes of the problem” and provides a comprehensive understanding of the nature of the problem, its manifestations, determinants, or factors that cause the problem, the conditions under which the problem comes about and may change over time, and the relevant consequences of the problem. Understanding these characteristics of the problem is critical for generating intervention strategies that are consistent with and responsive to them. For instance, interventions targeting the initiation of behavioral change differ from those targeting behavioral maintenance. Variations in the characteristics of the problem require interventions of different nature and with different goals and structure. Thus, before we develop the intervention, we need to specify the characteristics of the problem that the intervention is designed to address. Failure to clearly outline the nature of the problem beyond its behavioral manifestation limits our ability to tie the problem clearly to an intervention. In a review of physical activity interventions, Keller, Fleury, Sidani, and Ainsworth (2009) note that few of the studies reviewed specified the problem under investigation from a theoretical,
rather than behavioral, perspective. Many intervention studies outline intervention components, but do not clearly specify the problem to which the intervention is directed. Instead, the focus of the research is on specific behaviors, disease states, or health-related outcomes, inferring that outcomes of the intervention reflect the problem addressed by the intervention.

This may not always be the case; the problem to be targeted may be related to a determinant of a behavior or health state (e.g., lack of motivation to engage in physical activity) or the conditions under which the problem comes about (e.g., unsafe neighborhood). Although specifying the nature of the problem is helpful as a beginning step in problem specification, additional exploration of the problem characteristics is required to identify which is amenable to change and, therefore, can be meaningfully addressed by the intervention. Problem definition goes beyond the recognition of disease processes, such as obesity, diabetes, or cardiovascular disease, to detail its essential attributes or experiences such as failure to maintain dietary changes consistent with limited self-monitoring skills or lack of self-efficacy to engage in physical activity and its determinants or causative factors. Slippage between the specification of the problem and the intervention increases error in the design of the intervention, which results in limited, if any, effectiveness of the intervention in addressing the problem, and reduces our ability to make valid causal inferences. Even when interventions are effective, if the problem is not clearly specified, the ability to replicate intervention findings is limited.

Our emphasis on understanding the problem in intervention research is to call attention to the need to carefully design interventions that are responsive to an identified problem. Understanding the problem requires clarification of the nature of the problem, its manifestations, causative factors, and level of severity, and using this information to identify a theory that explains these characteristics of the problem (see Exhibit 4.1). Specificity in understanding the problem may come from qualitative and descriptive data that identify the relevant dimensions of the problem, when the problem occurs, as well as how the problem may look or come about differently given different social conditions and populations. Descriptive data may provide important information about problem manifestations, particularly given differing social conditions and different populations, the level of problem severity, and the most significant determinants of the problem such as perceptions, beliefs, personal enabling skills, and social enabling factors that should be taken into account when developing an intervention (Thornton et al., 2017).

Qualitative methods have been considered necessary for developing valid and valued interventions that acknowledge (a) cultural and contextual factors that facilitate the effectiveness of intervention, (b) social and ecological validity of intervention, and (c) attention to the specific needs and resources of the target population (Duggleby & Williams, 2016). Similarly, epidemiological and community-based data can provide direction for identifying concrete problems in communities. Descriptive and qualitative data may be helpful in identifying individual and community experiences in problems of living or taking people’s problems as a starting point to determine if intervention makes a difference to those problems in practice. Our understanding of the problem allows us to address the manifestations of the problem through questions such as “How would you know the problem if you saw it?” Manifestations of the problem typically include patterns of signs and symptoms experienced in association with the problem. Specificity in understanding manifestations of the problem may come from descriptive or clinical data that specify the relevant dimensions, when the problem occurs as well as how the problem may present differently given different social conditions and populations. It is important to frame understanding of the manifestations of the problem as it relates to a specific condition, population, or behavior rather than assuming generalizability across conditions. In a critique of the transtheoretical model (TTM), Hutchison, Breckon, and Johnston (2009) note that although the TTM has been widely applied in interventions designed to promote physical activity, no attempts have been made to ascertain whether different change processes occur in initiating physical activity behavioral change beyond those specified by the TTM. Thus, different processes of change, or manifestations
of the problem of limited motivation to engage in behavioral change, may exist for sedentary individuals compared to smokers or those with a poor diet.

There are typically multiple determinants or causative factors for a problem. These are likely to differ in scope, with some focused on individual function, resources, or cognitions as the source of the problem, whereas others focus on community, organizational, or policy factors. For example, possible determinants of physical inactivity may include diminished functional status, lack of social support, perception of an unsafe environment for activity, or public policy, which limits the construction of walking trails or neighborhood parks (Thornton et al., 2017). By carefully considering the determinants or causative factors associated with a problem beyond its behavioral manifestation, we are better able to target our intervention strategies to address the specific factor and subsequently the problem in a responsive and systematic way.

The problem also is characterized by its intensity, severity, and duration. Understanding these characteristics involves the identification of patterns such as the presence and perceived burden of symptoms that accompany a particular problem; changes in psychosocial and functional status across the trajectory of chronic illness; or perceived capability, opportunity, and motivation related to problem maintenance. Our understanding of the level of problem severity may be addressed through questions such as, Where and when can change be expected and which outcomes are central? Attention to these characteristics helps to ensure that the problem is addressed in ways that will produce the desired effects with interventions of appropriate nature and/or dose. For instance, recommendations for diet and physical activity will differ for people with varying levels of weight excess.

Considering when the problem exists or when the risk of the problem is greatest may help to understand variations in problem intensity and frequency that take place naturally over time or that are associated with particular conditions. For example, descriptive data on the maintenance of physical activity following cardiac rehabilitation shows that the risk of relapse is greatest within 2 weeks following completion of a monitored program. Specification of the trajectory of the problem allows us to localize the most effective time for intervention delivery through questions such as, How does the problem change over time, place, events, or people? Determining the appropriate time for intervention allows us to target the intervention delivery to when the problem achieves a level of sufficient intensity to require intervention. When timing is not considered, interventions may be given without considering when the intervention is most likely to produce the greatest effect, potentially leading to incorrect conclusions about its effectiveness. Clinical and descriptive data, particularly those derived from longitudinal observations, may guide determination of the most appropriate time for intervention delivery.

Desired outcomes reflect both achievable and clinically meaningful changes in the problem. However, our understanding of the problem guides the specification of outcomes. Expected outcomes will differ in nature, timing of when the changes are expected to occur, and the expected pattern of change after implementation of an intervention.

The nature and characteristics of the problem can be integrated in a theory of the problem. It models the processes that produce the problem needing attention. Problem theory may be judged according to how clearly it addresses the phenomenon of concern or explains the problem requiring intervention. It provides a framework for deciding what characteristic of the problem is treatable, when and at what level it is treatable, for what populations, under what conditions, and for what desired outcomes. The problem theory gives directions for delineating intervention strategies or ways that will successfully address the problem and produce the desired outcomes. Knowledge of the conditions under which the intervention is expected to work is necessary for generalization of its effects, for improvement or refinements of the intervention, and for clinical applicability.

The problem theory offers an appropriate grounding for stating the research topic to be investigated in an intervention evaluation study, for identifying the study target population and setting, for delineating the participants’ eligibility criteria, and for justifying the timing
for the delivery of the intervention. Specifically, adequate definitions of the problem rely on a specification of the nature of the problem and its characteristics as experienced by the target population. Particular attention to the degree to which the conceptualization of the problem reflects the experience of the target population is critical. Such descriptions, tied to clear conceptualization of the problem, are needed for selecting appropriate samples and assessing the applicability of findings to other populations. Often, it is assumed that there is a shared meaning for constructs. However, culture determines perceptual, explanatory, and behavioral responses to health, health problems, and treatment. In order for a theory to adequately represent the problem in a specific population, the main concepts in the theory must be recognizable to the population and reflect identifiable events in their lives.

The adequacy of problem theory may be judged according to how clearly it explains the problem under investigation. In selecting a problem theory, it might be helpful to explore available theories that have been generated to explain the problem. Typically, theories are chosen that have been used successfully in previous pertinent research. However, evaluation of the degree to which these theories reflect the experience of the population of interest is needed, particularly among vulnerable groups that have traditionally been absent from clinical studies. A challenge for researchers and clinicians is to identify, develop, or refine the multiple theories, or derive the problem theory through (a) conceptual analysis of available theoretical, empirical, and clinical literature resulting in the generation of a meaningful and coherent combination of various factors within a broad framework that comprehensively explain the problem; (b) integrating empirical, quantitative, and/or qualitative evidence as is done in literature review or meta-analysis to clarify the problem; and (c) carefully designing and carrying out descriptive correlational studies that serve as early exploratory efforts to establish the characteristics of the problem, encompassing its nature, severity of its presentation, manifestations, and determinants or causative factors, in different populations and under different conditions.

Specifying the Population Experiencing the Problem

Conceptualization of the problem targeted by the intervention provides relevant information for specifying participants’ characteristics or attributes that might determine response to the intervention, including the nature, course, and manifestation of the problem. Effective intervention planning is based on how the problem is experienced by the target population and how the proposed intervention will address the problem in that population. Knowledge of “for whom, and under what circumstances” the intervention works is necessary to maximize treatment effectiveness. A systematic evaluation of intervention effectiveness requires that the intervention be given to individuals who are experiencing the problem or its determinants that are treatable by the intervention. Participants who do not have the problem of interest cannot be helped by treatment, no matter how well the intervention is designed. In this case, results of evaluation indicate the interventions are not effective in addressing the problem, a potentially incorrect conclusion. Participant homogeneity in response to treatment is necessary to detect significant treatment effects in an experimental study or randomized clinical trial. In contrast, heterogeneity will produce variability in response to treatment reflected in increased variance in outcomes. Careful sample selection is needed to enhance homogeneity. Sample selection proceeds from our understanding of who has the problem or relevant determinants that are targeted by the intervention. Inclusion criteria reflect characteristics of the problem that are addressed by the intervention, meaning that we select those individuals who experience the problem as delineated in the problem theory because they are likely to respond to the intervention. In some instances, the conceptualization of the problem indicates the condition under which the problem occurs; that is, it will tell us that the problem is inherent in a given situation. For example, lack of resources for physical activity or an unsafe environment might be inherent to residents of a given neighborhood; in other situations we may not know when the problem will
occur exactly. For example, social cognitive theory does not guide us in knowing when a lack of self-efficacy might occur, so it is necessary to assess these perceptions as part of inclusion criteria. Characteristics of the problem must be measureable in some way to ensure that the sample is homogenous on what matters. A lack of specificity in problem identification creates the risk of delivering a theoretically relevant intervention to those who may not need the treatment. For example, Allison and Keller (2004) tested an intervention based on social cognitive theory and designed to promote maintenance of physical activity among older persons following graduation from cardiac rehabilitation. It was proposed that risk for relapse was due to low levels of efficacy expectations; levels of efficacy were not measured as part of study inclusion criteria. Failure to show significant changes in self-efficacy following the intervention was because of participants having high scores on measures of self-efficacy at baseline, creating a ceiling effect. Thus, in this case, the problem was poorly conceptualized, the sample selection was not well specified, and the intervention was delivered to those who did not need it.

To build intervention science and to facilitate knowledge translation, we need to improve our understanding about how interventions function as participant characteristics vary. The major chronic illnesses—heart disease, diabetes, stroke, and cancers—are disproportionately common among vulnerable populations defined as those who live in poverty, have limited education, live in rural areas with limited access to healthcare, and the very old. Such populations may respond differently to interventions that were successful in the majority populations. Thus, there is a need to know whether interventions have differential effects upon participants with varying characteristics of vulnerability. This, in turn, will increase our understanding of the processes through which the intervention operates with certain participants and not with others. Identifying the responsiveness of different groups to different or same interventions can yield important information about the benefits of the intervention for at-risk individuals. It is likely that different people change their behavior for different reasons. A key will be to determine what interventions work with what groups of people and understand when to select the optimal time frame for intervention.

Attention in intervention research has recently shifted to the development and evaluation of tailored and stage-matched interventions, which require a specific focus on sample selection and intervention delivery. In tailored interventions, specific levels or types of treatment are matched to specific participants’ characteristics assessed at pretreatment, using validated algorithms. This design may improve the efficiency and cost-effectiveness of interventions by delivering treatment in a way that is consistent with and responsive to individuals’ profiles. Depending on the focus of the intervention, tailoring variables may include individual, family, or context characteristics. Collins, Murphy, and Bierman (2004) introduced the use of adaptive interventions as a perspective on research-based prevention and treatment. An adaptive intervention assigns different doses of certain intervention components across individuals and/or within individuals across time. More recent applications include Just-in-Time Adaptive Interventions (JITAIIs) designed to address the changing needs of individuals by providing the type and amount of intervention support needed, when it is needed. Advances in mobile technology have fostered real-time, individualized approaches responsive to the intervention needs of participants, with doses assigned on the basis of decision rules linking characteristics of the individual with specific levels and types of intervention components. Articulating the links between doses and types of treatment and participant or problem characteristics optimizes intervention effects. Prior research may be helpful in specifying these relationships, as well relevant theory and clinical experience.

Identifying Mediating Processes
Mediating processes specify the mechanism underlying the intervention effects as a cause–effect sequence between an intervention and an ultimate outcome (Teixeira et al., 2015). The sequence states how an intervention leads to desired changes in the ultimate
outcome, clarifying the essential process of change beyond a simple input–output model. Few interventions can be viewed as working directly on the ultimate outcome; nearly all have their effects through some mediating variables such as reactions to the intervention, knowledge, self-efficacy, or perception of social norms. These mediator variables are critical cognitions, behaviors, or attitudes that must be changed by the intervention in order for the desired outcomes to occur (see Exhibit 4.1). The change process is contingent upon accurate specification and manipulation of mediating variables; interventions are more likely to be effective if the mediating variables are strongly related to the desired ultimate outcome and if the intervention clearly targets change in the mediating variables at acceptable levels. The potential effect of an intervention is characterized by the relationship between mediators and specified ultimate outcome, so specifying the mediating variables helps to clarify the change process.

Theory should clearly specify relevant mediators. The mediators chosen need to be consistent with and reflect the mechanisms for change connecting the intervention or its components with the intended ultimate outcome. Further, mediators must be amenable to change by the intervention, and the degree of change in the mediators following treatment delivery must be measured in a standardized way. Accordingly, theory guides the selection of instruments for measuring the proposed mediators, the delineation of the time at which and the pattern of changes in mediators are expected, and the plan for data analysis. The analysis is concerned with examining the extent to which the intervention has direct effects on the mediating variables and indirect effects on the desired ultimate outcome. The links will clarify the mechanisms for change and allow a greater understanding of how intervention strategies are expected to lead to desired outcomes. Knowledge of why and how the intervention works is necessary for its careful application in practice.

**Developing the Intervention to Address the Problem**

Theory offers explanations of the problem amenable to treatment, indicates the nature of relevant intervention strategies that could successfully address the problem, as well as the processes mediating the intervention effects on the ultimate outcome. The design of theory-based intervention involves feedback or movement between the conceptualization of the problem and the specification of key elements of the intervention designed to treat or prevent the problem. Thus, the intervention is developed in the context of the problem and is specifically developed to address the deficits identified in or to promote the strengths inherent to the problem. Determining what makes the problem better or worse helps to clarify the specific components of the intervention, or identify the related antecedent or associative factors. Together, they specify the structure and content of the intervention, including critical inputs, components, treatment strength, and intervention dose and timing. As such, they provide a basis for explaining how an intervention exerts its influence on outcomes and further understanding of what interventions work, for whom, and under what conditions. When these elements are considered, the possibility of alternative rival hypotheses about the intervention’s contribution to the outcomes is minimized, thereby strengthening internal validity (i.e., the degree to which we can say that the change in the outcome is due to the intervention, not extraneous/confounding variables).

The theory guides the operationalization of the intervention; it specifies the nature and dose of the intervention necessary to produce the expected effects. The intervention is operationally defined according to the central concepts of the theory as the critical inputs that define the intervention. Intervention science cannot be advanced without clear identification of the specific actions and the conditions under which an intervention is delivered. The intervention is operationally defined in terms of its “active ingredients” or specific elements that are responsible for creating change in mediating variables and/or outcomes. The active ingredients are activities or strategies that constitute the intervention; they represent what is
to be done and what is necessary, sufficient, and optimal to produce the desired outcomes. In addition, the theory specifies the mode for intervention delivery and the strength and dose required to bring about desired change in the outcome in the target population. This level of specificity enables identification of what comprises treatment and what does not, thus clarifying the critical aspects of the intervention and the possibility for variation in treatment delivery, minimizing the possibility of alternative or rival hypotheses, and strengthening internal validity. To achieve specificity, the essential elements or features of the intervention are identified and separated from the less important ones. This is important to (a) distinguish the intervention from others (this will guide the selection of an appropriate comparison treatment, which should not incorporate the theoretically active ingredients of the intervention under evaluation), and (b) develop instruments to assess and monitor fidelity of intervention implementation by interventionists and adherence to treatment by participants. Keller et al. (2009) found that despite differing theoretical perspectives guiding interventions to promote physical activity, studies reviewed incorporated similar intervention approaches, designed to enhance cognitive and behavioral processes of change, decisional balance, self-efficacy, social support, self-regulation, and outcome expectancy. The similarity of intervention approaches across studies reviewed blurs the theoretical specificity of treatments and limits our ability to clearly evaluate the efficacy of one perspective over another. A greater level of specificity in operationalization of interventions that are consistent with relevant theory is needed to maintain construct validity of the intervention implementation. The current state of intervention research has provided little guidance on how to best manipulate mediating variables to obtain behavioral change. Greater theoretical specificity in developing critical inputs that target theoretical mediators may address this concern (Teixeira et al., 2015). Similarly, specificity in how constructs are operationalized will foster the effective blending of different behavioral theories as a guide for intervention, recognizing how each contributes uniquely to the intervention and specified outcomes.

Treatment strength refers to the likelihood that an intervention can achieve the desired ultimate outcome. Investigators use information from theory, descriptive research, and prior intervention to propose the amount or format for intervention delivery needed to produce a measurable effect. Strong treatments contain large amounts of the “active ingredients,” which lead to change in the outcomes. If treatment strength is low relative to the problem of interest, it is unlikely that significant change will be achieved. Although evaluation of intervention strength is viewed as essential in pharmaceutical treatments, where the dose continuum is largely a matter of quantity and frequency, few theory-based interventions in the behavioral and social sciences have evaluated treatment strength.

Treatment strength may be characterized as including several factors central to intervention design and testing, including clarity in intervention conceptualization and operationalization, treatment dose, and fidelity in treatment implementation. Treatment strength is enhanced by a clear theoretical rationale for intervention design, with specified links that tie the intervention’s critical inputs to mediating processes and outcomes. A key consideration is the conceptual relevance of the treatment or specification of what treatments will address which problems; if a treatment does not address the problem of interest, strength is irrelevant. To evaluate treatment strength, we must be able to identify what treatment addresses which problem of interest, as well as what constitutes treatment and what is not consistent with theory, and where treatment delivery might vary without jeopardizing its integrity and effectiveness. This level of conceptual specificity will foster the development and use of detailed treatment manuals and protocol to guide interventionist training, determine intervention dose and delivery, and evaluate treatment fidelity. A standardized protocol adds clarity and facilitates intervention delivery as designed, which keeps it from becoming contaminated or weakened over time. Standardization may include developing a protocol that describes the nature of the intervention activities, specifies the sequence of activities to be performed, and
provides details for the procedures to be carried out (Sidani & Braden, 2011). Procedures for quantifying the strength of intervention may include a quantitative synthesis of past research to summarize components of “strong” interventions, expert review and evaluation of intervention strength, and programmatic research to systematically test intervention of different strengths. Methods to ensure treatment strength include the use of a scripted protocol, training of the interventionist, a structured review of information shared with participants, and intervention delivery audit and monitoring.

The dose of an intervention represents the amount, frequency, and duration with which the treatment is given to produce changes in mediating and/or outcome variables. Response to treatment is often characterized as a dose–response relationship. Amount refers to the quantity of the treatment that should be given. Frequency refers to the number of times the treatment is to be given over a specified time. Duration refers to the total length of time the treatment is to be implemented for the expected changes in outcomes to be achieved. Intervention dose may be characterized in different ways, depending on the focus and mechanisms for intervention delivery. A “dose” of an intervention might be a telephone call, session attendance, the number of contacts or minutes engaged with a lay health advisor, completion of an online educational module, or number of postings or minutes engaged in an online support group. Dose might be carried out weekly, biweekly, or monthly and continue over 6 weeks, 8 weeks, or 6 months. The optimal dose of the intervention needed to achieve effects may be determined through clinical experience, programmatic research on the intervention, or an integrative review of intervention studies. One approach to determining dose may be to test interventions in which different dose levels are implemented and compared. The process of quantifying intervention dose includes developing a measure assessing intervention amount, frequency, and duration that reflects the dose received by participants. Intervention dose may be quantified by sessions attended or telephone calls received; a log in which participants document what they did, when, how frequently, and for what length of time. Quantifying intervention dose allows for a dose–response analysis in evaluation of intervention effects, which assists in identifying or confirming the optimal dose for producing relevant outcomes (Conn & Chan, 2016).

Treatment fidelity refers to the methodological strategies used to monitor and enhance the reliability and validity of intervention conceptualization and implementation. When there is a high degree of monitoring and control over factors associated with the delivery of intervention, it is possible to evaluate the efficacy of a theory-based intervention, or compare the impact of two or more treatments on an outcome with validity. Unless treatment is implemented with fidelity, the extent to which the intervention under evaluation is the primary mechanism for the observed changes in the outcomes will remain unclear. Further, intervention strength is dependent on treatment fidelity, with low fidelity decreasing the strength of the intervention. Bellg et al. (2004) outline five areas to conceptualize and maintain treatment fidelity: (a) study design, (b) training providers, (c) delivery of treatment, (d) receipt of treatment, and (e) enactment of treatment skills.

Treatment fidelity related to study design includes a strong and identifiable integration of theory within the conceptualization of the problem targeted by the intervention, operationalization of the intervention, specification of mediating processes, and specification of the ultimate outcome (Ibrahim & Sidani, 2015). Approaches to maintain treatment fidelity specific to study design are intended to ensure that the intervention is consistent with the underlying theory, allowing an accurate test of theory-based interventions. The effect of an intervention can be accurately evaluated when the research design does not confound treatment effects with extraneous differences between treatment groups or treatment and comparison groups. Treatment fidelity is strengthened with the development and use of a standardized manual and protocol for implementing the intervention as conceptualized. Careful delivery of the intervention as specified in the manual or protocol decreases the potential for contamination across treatment
groups. To date, intervention research targeting behavioral change has shown variable fidelity to theory in study design, limiting judgments about the contribution and explanatory power of health behavioral theories, informed comparison among theories, and systematic intervention replication (Keller et al., 2009).

Training of the interventionists is an important area to consider in maintaining fidelity. Variability in intervention delivery will make it less likely that significant effects will be achieved and will introduce alternative explanations for intervention effects. Treatment fidelity is enhanced with standardized interventionist training in the theory underlying the problem and the intervention, the protocol for intervention delivery, and any procedures specific to participant contact or follow-up. Standardized training of interventionists to performance criteria is continuous throughout the study period. Methods of training may include standardized training materials, conducting role-playing, and observing intervention delivery for consistency with the manualized protocol. The risk for slippage in the implementation of the intervention over time may be addressed by interventionist-training booster sessions, regular review of session audio or videotapes for consistency with the manualized protocol, and review of interventionist documentation of intervention delivery. The adequacy of training in both theory and intervention implementation is evaluated and monitored for each interventionist both during and at regular intervals after the training process.

Treatment fidelity processes monitor and improve implementation of the intervention so that it is delivered as intended (Moore et al., 2015). In testing theory-based interventions, significant treatment effects are observed in the form of differences between the outcomes for the intervention group and those for the comparison group. Thus, the clear specification of the content and activities constituting the intervention and the comparison treatment conditions and strategies to support their consistent delivery is essential to enhance internal validity. A primary challenge to treatment fidelity is variability in treatment implementation either across interventionists and participants and/or over time. If the delivery of intervention changes over time, or if it is inconsistent because of different interventionists, across participants in different sites or at different time points over the study period, then variability in participant response to treatment ensues; this, in turn, decreases the likelihood of finding significant intervention effects. Indeed, process evaluation of a randomized trial to promote self-management in primary care patients showing no significant effects (Kennedy et al., 2013) revealed that the intervention as designed was not delivered to most patients (Kennedy et al., 2014). Another challenge is limited contrast or difference in treatment delivered to the intervention and comparison groups. The difference between groups, specific to the theoretically active ingredients of the intervention, represents the strength of the intervention effects on the desired outcomes. Treatment conditions without clearly specified content and clear differentiation will reduce any differences found in outcomes between the intervention and comparison groups. In addition, monitoring intervention delivery requires strategies for quantifying the dose proposed to produce the desired change in outcomes. Abildgaard, Saksvik, and Nielsen (2016) outline qualitative and quantitative approaches to evaluating the intervention process. Quantitative approaches include the use of process evaluation scales measuring key aspects of the intervention and degree of implementation (Havermans et al., 2016). Process evaluation scales document the extent to which the intervention content was relayed, activities were performed, and dose was offered in accordance with those specified in the treatment manual and protocol. The use of validated process evaluation scales during intervention delivery may allow investigators to better understand factors that limit or promote implementation of the intervention, such as a lack of time or group dynamics, and better inform future research. Process evaluation scales completed by the interventionist following each session may highlight areas of weakness or alterations in treatment delivery and reinforce the need to address these issues with the interventionists. Randomly audio- or videotaping treatment sessions may
be helpful as a post hoc evaluation of treatment delivery. Comparison of what was delivered to what was specified in the treatment manual and protocol identifies discrepancies that should be accounted for at the data analysis stage. Regular review of process evaluation data and feedback to interventionists may strengthen the skills for providing the treatment and support the standardization of intervention delivery across interventionists, participants, and sites. A qualitative approach to process evaluation includes collecting and analyzing data through interviews, focus groups, field notes, or observation. Qualitative approaches to process evaluation may help to explain unexpected results of an intervention as well as clarify intervention mechanisms (Greasley & Edwards, 2015).

According to Bellg et al. (2004), receipt of treatment involves processes that monitor and improve the ability of participants to understand and perform the necessary behavioral skills and cognitive strategies central to treatment delivery. Treatment receipt evaluates participant use of behavioral skills (physical activity monitoring) or cognitive strategies (self-regulation) that have been developed as part of the intervention. Methods for evaluation include review of homework assignments, documenting strategies used to achieve personal goals, or review of physical activity calendars.

Enactment of treatment consists of processes to monitor and improve the ability of participants to perform the necessary behavioral skills and cognitive strategies in relevant real-life settings (Bellg et al., 2004). Enactment evaluates the extent to which participants use specific behavioral skills or cognitive strategies at the appropriate time and setting in daily life. Enactment evaluation may include self-report of behavior over time, use of monitoring devices such as accelerometers to evaluate physical activity patterns at home, or telephone calls to monitor the use of coping strategies.

Careful consideration of the timing and conditions for implementing the intervention contributes to consistency in treatment delivery and to internal validity. Timing refers to providing the intervention when it will be most responsive to the problem and most effective in creating change. Similar to biological treatments designed to prevent or treat illness, behavioral interventions require sensitivity to a timing factor. When timing is not considered, interventions are given without considering when the intervention is most likely to produce the greatest effect.

The conditions for delivery of an intervention should be consistent with those specified in the intervention theory. To rule out alternative explanations for intervention effects, it is important to distinguish the “active ingredients” of the intervention from the elements associated with its delivery methods and context. Issues in treatment implementation are aspects of the delivery system relevant to providing the specific intervention. These issues refer to the resources needed to carry out the intervention, including material supplies and human skills that facilitate the delivery of the intervention as planned. Intervention delivery targeting individuals or groups, at one or multiple sessions, may differ by population and setting and will reflect how the intervention is packaged for a given study (format of information, color, language, pictures used), the delivery plan developed (information through audiotape delivery, Internet delivery, mail delivery), and the identification of settings that facilitate the implementation of the intervention as designed. Understanding of conditions may come from analysis of the context under which successful interventions have been delivered, acknowledgement of conditions guiding descriptive or pilot work, clinical knowledge of the importance of settings, the timing of intervention delivery, and the role of supportive or unsupportive situations in resolving the problem. For example, an understanding of community ecology can lead to a better match with community-based health promotion interventions, including developing new and existing leadership, strengthening community organizations, and furthering organizational collaboration (Fleury & Lee, 2006). Thus, a thorough understanding of the multiple conditions under which interventions might best be delivered, including attention to intervention feasibility and cost, has the potential to facilitate the identification of interventions that are effective.
Context includes characteristics of participants and interventionists, and environmental factors that may influence intervention implementation. Relevant participants’ characteristics may include demographic profile (age, gender, education level), personality traits (locus of control), or personal beliefs (cultural values). Characteristics of the interventionist may include personal and professional attributes necessary for treatment delivery such as communication skill, educational background, beliefs about the treatment, or competence in treatment delivery. Environmental factors encompass physical and psychosocial features of the environment; they relate to the convenience of the setting in which treatment is given to participants; a setting that is clean, safe, private, and quiet; availability of resources necessary for treatment delivery; and the organizational culture of the setting reflecting prevailing norms and policies.

**Clarifying the Outcomes of the Intervention**

In contrast to a black box approach to intervention design and evaluation, the application of theory allows a more comprehensive modeling and measurement of the processes accounting for the ultimate outcome expected as a result of the interventions. The outcomes reflect the resolution or successful management of the problem targeted by the intervention. They are the reason for which the intervention is given and form the criteria for evaluating intervention effectiveness. Thus, outcomes selected are consistent with the target problem and responsive to the intervention. Theory determines (a) the nature of the anticipated outcomes, including clinical end points, functional status, or changes in knowledge, attitudes, and motivation; (b) the points in time at which changes in outcomes are expected to occur; and (c) the expected pattern of change in outcomes after delivery of the intervention. The nature of the outcomes refers to aspects of the problem that the intervention is designed to address and the ultimate goals of the intervention. By specifying the logic that connects intervention critical inputs to mediating and outcome variables, theory identifies the outcomes that can reasonably be expected. Without this specification, there is not a clear basis for selection of meaningful outcomes that are responsive to the intervention and of instruments to validly measure outcomes. The timing refers to the point at which the changes in the outcomes are expected to take place following intervention delivery. Some changes occur during or immediately after intervention delivery, whereas others appear over time. The pattern of change refers to the trajectory, or the direction and rate of change in the outcomes over time. The theoretical perspective, clinical experience, and descriptive data can provide direction for when to best measure change; change may be targeted on the basis of theoretically important times, at multiple time points to investigate trends over time, or at clinically important times. Many studies measure outcomes immediately following the treatment, or posttest, with limited justification for the timing of measurement. This approach to outcome measurement implies that effects have an immediate onset and peak following intervention delivery, but other patterns of change are possible. Different outcomes might have different patterns of change, with some most pronounced immediately following the intervention and others at some later time. Choosing appropriate times for outcome measurement is required to detect important intervention effects; otherwise, we may be looking in the wrong place at the wrong time.

**SOURCES FOR GENERATING THEORY**

There are several sources of theory available to guide the design interventions. Donaldson and Lipsey (2006) note that the choice of theory requires clarifying assumptions about the etiology of the problem the intervention attempts to address and the mechanisms by which
change can be achieved. Three approaches may be helpful in providing an explanation of the problem and the mechanisms by which change may be achieved.

Middle-range theories, which may be derived from relevant grand theory, can guide intervention development and testing focused on a limited aspect of experience in a particular situation of interest. A middle-range theory has a focus on specific health experiences, health and illness problems, or certain patient populations; it provides clearly articulated concepts and relationships, which can be tested empirically through systematic research. Theory may be developed through qualitative research and practice observations or through logical analysis and synthesis of existing theory (Liehr & Smith, 2017). Middle-range predictive theory allows the prediction of relationships between concepts, or how changes in a phenomenon occur. Shearer (2011) describes the iterative process guiding the development and testing of her middle-range theory of health empowerment, leading to a theory-based health empowerment intervention. Beginning with qualitative methods, Shearer identified aspects of the lived experience of health empowerment, including contextual factors and interpersonal factors positively associated with health empowerment. In quantitative testing of this emerging theory, Shearer and Reed (2004) examined the relationships among contextual factors, interpersonal factors, and purposeful participation in health behaviors. Although the study findings partially supported the theoretical propositions, the limited variance in behaviors explained by contextual and interpersonal factors warranted further study. A second qualitative study with older women clarified the social support needs and resources and contextual factors used to facilitate health empowerment (Shearer & Fleury, 2006). Social-contextual resources in the revised theory reflected perceived supportive relationships and opportunities for nurturance and the exchange of information and materials to foster health empowerment. A third qualitative study with homebound older women (Shearer, 2009) clarified the role of personal resources as the unique characteristics of each woman, including inherent strength as self-capacity that promoted change and growth. Health empowerment was defined as a relational process emerging from a woman’s recognition of personal resources and social-contextual resources, which results in a transformation in awareness of and belief in ability to knowingly participate in the changes inherent in health and health outcomes. The health empowerment intervention was developed to enhance awareness of personal resources and social-contextual resources, to foster purposeful participation in the attainment of personally valued health goals, resulting in well-being. Elements of the intervention are self-capacity building, social network building, and building social service utilization. The interventionist works in concert with an older woman and engages in a participatory process in which the interventionist listens and encourages the older adult to talk, share, and enact her health goals (Shearer, 2009).

Meleis (2010) presents a middle-range theory of transition and outlines an empirically supported framework to articulate and characterize the relationships between theory concepts. The use of inductive and deductive reasoning guided evaluation of the utility of different concepts of the framework and identified additional emerging concepts. Through a series of five studies, the framework was able to specify (a) types and patterns of transitions; (b) properties of transition experiences; (c) transition conditions, including facilitators and inhibitors; (d) process indicators; (e) outcome indicators; and (f) intervention critical inputs. Understanding the properties and conditions inherent in the transition process guides understanding of the complexities of the problem, the development of intervention critical inputs consistent with the problem, mediating processes, and outcome variables, as well as the conditions under which theory-based interventions might be most effective.

Qualitative and mixed-methods research may provide opportunities to gather information required to understand the etiology of certain problems and the mechanisms by which change can be achieved as a basis for intervention design. The primary purpose of qualitative inquiry is to capture the meaning of phenomena and relationships among concepts.
as they occur naturally, and as reflected in thoughts, language, and behavior, from the perspective of the participants. Morse (2017) describes qualitative outcome analysis (QOA), a procedure designed to qualitatively identify intervention strategies and examine outcomes. QOA builds on a completed qualitative study, moving from an understanding of individual experiences to identifying and applying interventions. Using qualitative methods, the processes, stages, and phases that occur during the course of a given phenomenon are described to the extent that a theory is developed, and the factors that alter the course of the phenomenon are recorded. Through QOA, a series of steps are outlined, which begin with identification of the behavioral manifestations of the phenomenon, clarifying and documenting the dynamics of the problem, identifying strategies to modify the problem, evaluating the efficacy of the strategies developed, and modifying and reevaluating the intervention strategies used.

Sullivan-Bolyai, Bova, and Harper (2005) present qualitative description as a relevant method for assessing, developing, and refining interventions to serve vulnerable populations and reduce health disparities. The goal of qualitative description is descriptive and interpretive validity to provide interventions that (a) focus on factors that promote access and use of services, (b) are acceptable and understandable to those experiencing the problem of interest, and (c) are sensitive to the cultural context of those experiencing the problem. The inclusion of formative qualitative research before intervention development may establish the need for intervention, define key constructs in a culturally valid manner, foster understanding of the factors that influence the outcomes, and identify key resources related to the intervention.

Natasi et al. (2007) present a heuristic for research and intervention development that uses an iterative research-intervention process. The research process begins with formative data to test a proposed conceptual model developed on the basis of existing theory and research. Qualitative research methods are used to identify and define the concepts specific to a particular culture or context. Findings from the qualitative research are used to construct a modified model and develop assessment and intervention tools to test the model. Evaluation research involves the triangulation of qualitative and quantitative methods to examine acceptability, social validity, cultural specificity, integrity, and effectiveness of interventions. The repeated application of mixed methods across cultures, contexts, and populations can be used to develop a theory that reflects both universal and culturally specific concepts, as well as address local cultural and contextual needs.

Intervention design may proceed from an integration of concepts either proposed by related theories or consistently found to be related. Although integration of theories may result in stronger interventions, a clear rationale is needed for the added value of this approach in furthering understanding of the problem, or providing an innovative explanation of the mechanisms by which change can be achieved. The blending of theories in intervention design requires an understanding of the strengths and limitations of different theoretical perspectives, as well as the underlying assumptions of each theory and the congruence of assumptions across theories. Treatment fidelity in intervention design requires clear specification of the contributions of each theoretical perspective as critical elements of the intervention. Specificity in construct definition as well as clear links between intervention theory and the critical elements chosen for the intervention are needed to evaluate specific predictors from theory and provide meaningful guidelines for intervention development targeting these constructs. Specification of the unique contribution of each theoretical perspective to intervention design will further empirical support for the contribution of combined theory to conceptual understanding of theory-based interventions.

Published reports of individual studies that evaluate the effectiveness of interventions or of meta-analytic studies that synthesize the empirical evidence of intervention effects
may provide a guide for selecting the most relevant theory with empirical support. However, these reviews may offer a limited perspective on complex issues and often produce conflicting findings.

Given these approaches to theory identification and development, investigators are encouraged to consider what constitutes a relevant theory and how a theory might address a problem experienced by a given population (Exhibit 4.2). However, the literature on theory evaluation and application as a basis for intervention design and testing is very limited. Prochaska, Wright, and Velicer (2008) outline criteria for evaluating theory based on perspectives from the philosophy of science (Dubin, 1977; Kuhn, 1977). Theory is evaluated according to which concepts are considered as explaining the phenomenon of interest, including (a) conceptual clarity, or that concepts are well defined and can be measured reliably and validly; (b) consistency, or the articulation of a logical and unified set of relationships between concepts; and (c) parsimony, the statement of concepts and relationships in the simplest manner possible because complexity makes operationalization for intervention development and testing of the full theory difficult. When evaluating theory as a basis for intervention, the clarity and relevance of underlying assumptions, concept definitions, and construct validity are essential.

Theory is further evaluated according to support for how concepts in the theory are related, including (a) testability, or the extent to which concepts can be operationalized and measured, and propositions can be examined; (b) predictability, or empirical support for the theory in predicting future change, or patterns of expected correlations either at one point in time or longitudinally; and (c) explanatory, or evidence of causality, such as that found in randomized controlled trials testing theory-based interventions. When evaluating theory as a basis for intervention, the empirical support for theoretical propositions, explanatory power, and predictive ability related to the problem of interest will guide theory selection. Evaluation moves beyond the review of individual studies to include integrative review or meta-analysis; attention to testing of the complete theory is important in determining the contribution to theory-based interventions.

**EXHIBIT 4.2 Evaluating Theory Relevance**

- What theories have been used in prior research?
- Which contains well-defined concepts that are operationalized, explicit, and internally consistent?
- Is there evidence for shared meaning for concepts?
- Which has conceptual definitions consistent with underlying assumptions?
- Which explains the problem of interest in the least complex manner possible?
- What descriptive work has been done?
- What is the strength of empirical work?
- To what extent are the relevant theoretical concepts measured?
- To what extent has prior research targeted relevant theoretical concepts in intervention design?
- To what extent has prior research clarified the theoretical basis for intervention critical inputs?
- To what extent does the intervention change targeted theoretical concepts?
- Are theoretical claims congruent with evidence in explaining why change occurred and why it did not?
- How has prior research characterized participant inclusion criteria or tailored intervention to measureable relevant concepts?
- To what extent does the theory generate new questions and ideas and add to the knowledge base?
- To what extent does the theory generalize to other situations, places, and times?
- To what extent have concepts been demonstrated to mediate the effect of the intervention?
Theory provides an understanding of the problem that the intervention targets, the nature of the intervention, and the mechanisms underlying the anticipated improvement in outcomes.

Theory fosters a systematic approach to intervention development and implementation that allow us to move beyond a simplistic, outcomes-focused approach to examining the central processes underlying program effects.

Theory helps investigators to identify both outcomes and mediating processes that specify a cause and effect sequence between an intervention and outcome.

### REFERENCES


© Springer Publishing Company


©Springer Publishing Company


In intervention research, it is all about the outcome. Critical questions about measurement to ask yourself in preparing to conduct an intervention study are (a) What “so what” important outcome(s) do you want to change or influence with your intervention?; (b) How will you assess whether your intervention led to that change in outcome(s)? In intervention studies, you need to measure phenomena in a way that can then be analyzed statistically to determine whether your intervention made a difference in outcome(s) and, if it did so, was that difference beyond that expected by chance. You must measure variables of interest in a manner that allows you to determine true change in the dependent or outcome variable and what measures can accomplish this in a reliable and valid manner.

One of the first issues you must confront as an investigator is your choice of how detailed and specific your measures must be in order to adequately assess the constructs identified in your study. Every variable in your study must be defined, and its role within your model must be determined. These variables should be tied to key theoretical constructs and be measurable/observable. Before selecting your measures, you will need to operationalize your study variables (i.e., define how those characteristics in your sample can be measured and how they vary from subject to subject) and quantify them. These variables take two forms. Independent variables are those that precede or are antecedents to the dependent variable; they are also known as “predictor variables.” The dependent variable is the variable of most interest in intervention studies and is measured to assess impact of the treatment; it is referred to as the “outcome” or “criterion variable.”
One of the next decisions you will need to make is whether the variable should be measured using one of the four levels of measurement: (a) **categorical** variables, which are measured as unordered categories; (b) **ordinal variables**, which are measured as ordered categories with intervals that are not clearly equal; (c) **continuous variables**, which are ranked with quantifiable intervals; and (d) **ratio level variables**, which are continuous and have an absolute zero. Examples of categorical variables are one’s profession (e.g., teacher, nurse, psychologist, physical therapist, social worker) and race (African American, White, Asian). Gender is an example of a categorical variable with only two answer options (male/female), known as a “dichotomous variable.” Ordinal level variables might include degree of satisfaction (little, moderate, great) or perceived agreement with an issue (do not agree, agree somewhat, agree, strongly agree) and reflect a prioritization or leveling of answer responses, but the exact differences between each answer cannot be strictly quantified. Continuous variables do have both order and equally quantifiable distances between each category. Common examples include weight, pulse rate, milligrams of pain medication given or correct answers on an examination. Like continuous variables, ratio-level variables have both order and quantifiable intervals, but they also have an absolute zero point. Temperature is a common ratio-level variable. As scales progress from categorical to ordinal to interval levels, the measure of the construct becomes more precise. The end result of using scales with interval-level data is usually more ability to detect differences between groups in your intervention study. Even when categorical variables match well to the research question (e.g., studies examining whether an intervention leads to fewer patients diagnosed as obese), it is often best practice to measure the variable at a higher level, such as weight in pounds, and collapse the data later for analyses should you decide to do so. However, the collapse of data should be done with extreme caution as this consolidation of detail loses the very essence of differences between variable characteristics. Failure to quantify variables and quantify them at the highest level that is appropriate represents a lost opportunity to gather information that can help determine the impact of your intervention (see Table 10.1 for more examples of levels of measurement).

### TABLE 10.1 Determining Level of Measurement for Study Variables

<table>
<thead>
<tr>
<th></th>
<th>NOMINAL</th>
<th>ORDINAL</th>
<th>INTERVAL</th>
<th>RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable categories exclusive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Variable can be ordered/ranked</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Intervals between categories equal</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Absolute zero or absence of variable</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**LEVELS OF MEASUREMENT**

One of the temptations you will face in conducting an intervention study is the often overwhelming urge to include as many measures as possible so that you do not “miss” collecting data that could reveal the “scientific find of a lifetime” (many scientists have been kept awake at night worrying about this very issue). Although you may want to measure...
multiple variables to assess change in outcomes, it is critical to prioritize which outcome is most important—in other words, what you are “going to hang your hat on.” This being said, wise investigators often use multiple types of measures for their outcome (criterion) variable to ensure they can assess intervention efficacy or effectiveness while reigning in their use of multiple measures for every predictor (antecedent) variable. For example, in a study measuring the effect of a healthy lifestyles intervention on physical activity with teens, it would be wise to include both a self-report measure of activity along with an objective measure, such as data recorded by an actigraph of daily physical activity by the teens.

Clearly identifying your outcome variable is necessary to calculate power and determine sample size for your study. Conversely, not being selective with measures risks overwhelming your study participants, which can threaten your ability to collect data as you may risk losing several participants due to the burden you have placed upon them. You also need to determine measures for your study, balancing what is clinically important to assess with the actual logistics involved in each of the measures and the cost of the assessments. Therefore, measures used in intervention studies need to be (a) specific to the variable under study, (b) sensitive to change, (c) appropriate for the participants in the study, (d) cost effective, and (e) able to be collected in the minimum amount of time that the participant, researcher, and study sites can afford. Measures that are not psychometrically strong and lack evidence of validity or reliability place the investigator at risk for finding out absolutely nothing!

Many times interventions bring about very small changes in outcomes. Thus, outcome measures must collect data in a manner that can identify a small, but significant impact. In a HIV-prevention intervention targeted to adolescent girls, behavioral change would precede actual infection rates. Thus, pinning all hopes for intervention impact on measuring only HIV infection rates would minimize the impact of an intervention that successfully reduced sexual risk behaviors. When examining the appropriateness of measures, consider literacy levels, cultural tailoring, and developmental appropriateness. For example, studies of pediatric patients should not request long narrative responses to complex questions. The manner in which the measure is administered must be considered. Frail elderly participants asked to complete physically challenging measures might be unable to comply. Teenagers may be less likely to complete a diary requesting hourly information on symptoms or emotions. It may be difficult for participants to complete complex or open-ended questions in noisy waiting areas or public spaces. Measures also need to consume a reasonable cost within your budget. In addition, time and effort burden need to be considered for all persons involved in the study starting with those from whom you collect data to those collecting, coding, and analyzing them. Basically, you have to be thoughtful and judicious in your selection of measures in order to get the “best bang for your buck.”

The basic premise of measurement in research is delineated with this question: Are you measuring the outcome construct that is most important to measure and are you measuring it consistently? Measures should be objective and not prone to fluctuations based on factors that are influenced by the investigator (e.g., data collection personnel or instrumentation). The theory of psychometrics underlies the principles of measurement, and, thus, the assessment of reliability and validity of instruments used in intervention research is termed the “psychometric evaluation” of measures. Validity and reliability are really a matter of degree, not an “all or none” issue. Thus, as an investigator, you must provide evidence of validity and reliability, not “prove” that the measure is psychometrically sound.

**FORMS OF VALIDITY AND RELIABILITY**

Validity and reliability go hand in hand. Assessing validity of a measure, that is, gathering evidence to conclude that your measure actually taps the construct you intend it to measure, is one of the two forms of psychometric evaluation. Reliability is the ability of an
instrument to consistently measure what it claims to measure. A measure cannot be valid if it is not reliable; that is, it cannot actually assess a construct if the measure cannot do so in a stable dependable manner. However, a measure can be reliable without being valid because it can be consistently measuring something other than the variable you may think you are measuring. For example, you may have a self-report measure of depression with very high levels of internal consistency (i.e., reliability)—it certainly measures “a construct” consistently; however, it could actually be assessing something other than depression, such as anxiety.

How can you determine whether your measure is valid or not? There are some established criteria used to judge validity, including content validity, construct validity, and criterion-related validity (see Table 10.2).

**Content Validity**

Content validity refers to the actual “meat” of the measure; specifically, whether it reflects the body of knowledge surrounding the construct. For example, a measure assessing parent–child communication might address all forms of communication—verbal, nonverbal, written—rather than assessing only one of these facets. A measure of “test anxiety symptoms” would include both physiological and psychological symptoms. Investigators determine content validity early in the process of instrument development. The congruence between the conceptual and operational definitions of the variable is assessed in content validity.

There are two basic approaches for determining content validity. The most common approach to assessing content validity is to use experts in the field who confirm that the measure is tapping what it is supposed to measure and to have individuals who actually are experiencing the construct of interest to evaluate the overall appropriateness of the measure under question. For example, if you wanted to measure stress in school-aged children using a questionnaire, your first step in determining the measure’s validity would be to determine content validity by pulling together a group of “known experts” (e.g., school psychologists and nurses, counselors, and teachers) to review and critique the measure you have developed or selected. Their feedback on whether they believe the measure is assessing stress in school-aged children is used as one assessment of content validity. Interrater agreement between the experts on which items are pertinent can be used to provide evidence of content validity for the final items retained for the measure. Another approach to establish content validity is to ask individuals who “possess” the construct of interest (e.g., school-aged children experiencing high-stress situations or enrolled in stress-reduction programs) to provide feedback on whether the items on the measure reasonably assess or cover components of their experience. This approach of obtaining subjective feedback from

---

**TABLE 10.2 Forms of Validity and Reliability**

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>RELIABILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Validity</td>
<td>Test–Retest Reliability</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>Parallel Forms</td>
</tr>
<tr>
<td>Criterion-Related Validity</td>
<td>Inter-/Intrarater Reliability</td>
</tr>
<tr>
<td></td>
<td>Cronbach’s Alpha Coefficient</td>
</tr>
<tr>
<td></td>
<td>Split-Half Reliability</td>
</tr>
</tbody>
</table>

© Springer Publishing Company
participants is often referred to as “face validity” (Morrison-Beedy, Carey, Aronowitz, Mkandawire, & Dyne, 2002).

**Construct Validity**

A second form of construct validity assesses how well the measure meets the theoretical expectations of either converging or diverging with other related constructs. For example, if you are interested in measuring stress, you might also contend that individuals in your study who score high on your measure of stress would also score high on a measure of anxiety. Thus, scores on these two measures would be positively correlated and provide evidence of convergent validity. Similarly, participants who score high on your stress measure would, theoretically, score lower on your measure of calmness (which we hope at the time you are reading this chapter, you are also experiencing). Scores on these measures, stress and calmness, would be negatively correlated and serve as evidence of divergent validity. Both convergent and divergent validity are types of construct validity.

Another way to assess construct validity is through the statistical method of factor analysis. Factor analysis can be used to assess whether the items in a measure cluster into predetermined variable groupings. Usually, these groupings are theoretically driven. Exploratory factor analysis can be used to reduce and refine a group of items within a new measure. Factor analysis can also be used to confirm how multiple items conceptually “hang together”; that is, it can provide evidence for the items serving as a cohesive representation of the construct. For example, in developing a measure to assess HIV-related knowledge (Volpe, Nelson, Kraus, & Morrison-Beedy, 2007), 62 items were formulated based on subdomains identified as important—prevention strategies, testing and treatment, misconceptions, and transmission information. The goal was to develop an overall measure of HIV-related knowledge. Content validity was determined by experts in the field and at-risk populations. The expert feedback was used to reduce the items, based on redundancy or irrelevance. Confirmatory factor analysis revealed that the subscales reflected these predetermined areas, but the largest percentage of variance captured in the analysis supported that all of the items were relating to one large factor. The HIV Knowledge Questionnaire (HIVKQ) was construct validated using a statistically driven approach.

Validity can also be evaluated by comparing or contrasting measures on “known groups” with your intervention participants. For example, in the previously mentioned study, evidence for the validity of the measure was obtained by comparing total HIVKQ scores between groups (i.e., HIV experts, college students, and three community samples) with predictions made on which groups theoretically should have higher or lower HIVKQ scores (Volpe et al., 2007).

Construct validity can also be evaluated by conducting hypothesis testing. In this approach, you test assumptions about the variables in your study using hypotheses. If the testing supports this assumption, evidence for the construct validity of the measure is provided. For example, data were collected on the construct validity of the HIVKQ during intervention work that provided treatment-related evidence. We hypothesized that participants in a HIV-prevention program, which included an HIV-education component, would score higher on the HIVKQ than those in the control group. Mean scores on the HIVKQ were significantly higher for intervention versus control participants, thus providing evidence for construct validity of the measure (see Table 10.3).

**Criterion-Related Validity**

Finally, assessing whether a measure truly measures what you say you are measuring can be accomplished through criterion-related validity. This form of validity compares your selected measure with what is considered the “gold standard” for assessing that construct
of interest. Therefore, if you have developed a wonderful measure assessing depression (let us call it the “WMAD”), you might chose to assess whether scores on your measure correlate positively with a well-validated measure, such as the Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996). If your measure is a valid measure of depression, you would expect that participants who score high on the BDI would also score high on your wonderful new measure. Data can be collected at (a) one time using both your predictor measure (WMAD) and gold standard criterion measure (BDI) or (b) at two different time points where the criterion measure is administered at some time after the predictor measure. These are termed “concurrent” and “predictive validity,” respectively.

### RELIABILITY

When selecting measures for your intervention study, it is critical that they consistently assess your variable of interest. Assessing reliability can be thought of as determining the dependability or stability of your measure. Measures that are reliable are also considered accurate. In essence, any error within the measure itself is minimized. If you had a scale that at one time point gave you a weight of 150 pounds but a few minutes later gave you a reading of 145, you would not say the scale was reliable (although you might prefer the lower reading). Various forms of reliability measures are available that focus on the instrument’s stability, equivalence, and homogeneity.

### TESTS OF STABILITY

#### Test–Retest Reliability

Although there are several methods that can be used to assess an instrument’s reliability, you need to be selective and targeted in your choice(s). Assessing the stability or consistency of a measure from one time to another is derived through test–retest reliability. In this approach,
you would ask the same participants to respond to the same measure at a specified time interval (e.g., 4 weeks apart). Although there are often standard assessment periods used commonly by investigators (e.g., 2-week test–retest interval for self-report measures), the time period has to make sense to you as an investigator. Calculating a reliability coefficient based on comparing the scores at both time periods for each participant provides a statistical measure of the instrument’s stability.

Consider a situation in an intervention study where you might want to assess a participant’s knowledge of pregnancy prevention methods using your newly developed “Contraception Information Scale.” An important step in using your measure would be to pilot test it for usability, readability, completion time, and then assess its psychometric properties. For test–retest reliability, you might ask a group of women to complete your measure and record each woman’s score for an overall assessment of their knowledge. You may then ask these same participants to retake the measure using a predetermined time interval (e.g., 2 weeks apart), one that would ensure participants were not simply responding from memory yet be within a time period that did not allow for a change in knowledge due to intervention or historical occurrence. You would expect that, without some type of educational intervention factor, such as a visit to a healthcare provider for contraception, the scores on your measure of contraceptive knowledge would be relatively (not necessarily exactly) the same from time 1 to time 2.

### Equivalence

#### Inter- and Intrarater Reliability

The stability of a measure can also be assessed by comparing the scoring on instruments between at least two trained independent raters. In observational studies, this is referred to as interobserver or intrarater reliability, and the scores of the two independent raters–observers are compared for equivalence, with higher levels of agreement or correlations indicative of a measure that can be used consistently regardless of the person conducting the assessment. When rater scores are compared to assess percentage agreement, a target of 90% or greater agreement between them should be met in order to ensure equivalence and fidelity of this measurement within your study. For example, in an observational study of maternal–infant interaction, two independent raters observed mothers and infants interacting with each other using a 16-item dichotomous scale that measured positive and negative interactions. Each rater checked “yes” or “no” for each of the 16 items on the scale (e.g., mother speaks quietly to the infant; mother quiets the infant by rocking). The raters’ item-to-item responses were compared for equivalency, and the percentage of agreement was determined. Interrater reliability estimates comparability between different raters, whereas intrarater reliability assesses, in a similar manner, comparability or consistency in the same rater across multiple time points. Statistical results indicating 90% agreement or higher between or within raters provides evidence of the measure’s reliability. Relying on correlational analysis for this reliability rating could result in scores with high correlations, yet the raters do not actually agree on specific item scoring. Throughout an intervention study, intermittent reassessment of intrarater reliability should take place to assess rater drift in scoring. If drift has occurred, retraining of raters on the instrument is necessary until a 90% agreement is again reached. Reliability assessments also can take place within one rater conducting repeated scoring on the same measure; however, the items on the measure should be placed in random order for each completion.

#### Parallel Forms

Similarly, assessing the correlation between two versions of the same instrument is a measure of reliability using a coefficient of equivalence. Comparing parallel forms of a measure with one sample at one time point can provide evidence for the reliability of the measure.
To balance the impact of testing effects and fatigue, the measures should be administered in different order to participants; this approach is intended to reduce error rate. Oftentimes, the most common challenge encountered by investigators is that there are rarely true alternative forms of one measure available for use. These two instruments should have means and standard deviations that are approximately the same in order for the forms to be considered parallel. Acceptable correlations between the scores on both measures provide additional evidence for the reliability of the measure.

**Tests of Homogeneity**

**Cronbach’s Coefficient Alpha**

Another method of assessing the reliability of a measure with multiple items is to determine its internal consistency by calculating a statistic called the “Cronbach’s coefficient alpha.” This statistic is by far the most common method for assessing homogeneity in a measure. We can quantitatively describe the consistency of how items within a measure relate to each other, that is “hold together” as one construct, using this analysis. If items within an instrument are not all related, or homogeneous, internal consistency will be reduced. For example, although a measure of locus of control may address the overall construct, the items may be heterogeneous and measure two aspects of the construct, internal and external locus of control. This statistic can be calculated for instruments that obtain continuous- or interval-level data. The Kuder–Richardson formula (KR-20) is a similar statistic used in the case of dichotomous data. This form of reliability is influenced by two forms of potential variance: the number of items in the measure and the interitem correlations between items. Increasing the number of items that correlate with the rest of the items will increase the reliability coefficient. However, savvy investigators always recognize that these modifications also increase participant burden, completion, time, and cost. These data can be used to refine your measure to capitalize on increasing the measure’s reliability while balancing the aforementioned issues. Statistical packages will calculate a Cronbach’s alpha as well as each item-to-total and interitem correlation. Scores on both measures can range from 0.00 to 1.00, with higher scores indicating greater consistency between items. Conventionally, Cronbach’s alpha values of 0.70 or higher are considered satisfactory and acceptable as evidence of instrument reliability (Melnyk & Morrison-Beedy, 2012) although it is better to aim for using an instrument with an internal consistency reliability of 0.80 and above. A useful component of the analysis is the “alpha if item deleted” result. This function provides you with details regarding changes in alpha levels if each item in the scale was deleted. This strategy can help you to further refine your measure and eliminate redundancies. A word of caution in using this strategy—you, and not the computer, should be the judge on which items need to be retained even if the alpha might be increased slightly if certain items were removed.

**Split-Half Reliability**

Homogeneity within a measure can also be assessed using split-half technique. This method was used more commonly before computers made every investigator’s life easier and is simply taking one-half of the items in a measure and correlating the scores between the two halves. Oftentimes, items are selected by the “odd-even” rule (of course, measures with odd-numbered items present their own challenge). These two “half” scores should correlate very highly if the measure is consistent. Reliable measures with more items will have higher levels of consistency than briefer versions based on the formula used to calculate the statistic. Spearman–Brown can be used to correct the lower alphas obtained simply because of the reduced number of items in each comparison measure.
TYPES OF MEASURES

The scientific literature is filled with a plethora of instruments that can be used to measure your study’s variables. These can be classified into various categories which include, but are not limited to, self-report, biological, physiological, observational, and system measures. Knowing that you want appropriate, feasible, sensitive, and specific measures, the key is to be judicious in your choice of measures from an abundant “cornucopia” of choices. You will need to be selective and carry out a scientifically driven assessment of which instruments will best meet your needs for documenting change in your outcomes as well as antecedents. Evidence from pilot study data that supports the validity, reliability, and feasibility of your measures will do much to convince funding reviewers of their suitability for your intervention study.

The collection of self-reported data is extremely common in intervention research (e.g., ask the participants themselves if they liked the intervention, if they did the activities, if they felt stressed, if they have had angry outbursts, if they smoked a cigarette, if they love their mother). Much of intervention evaluation lies in these types of self-report responses. There are numerous mechanisms that can be used to collect this information; a few of the many options are presented here.

Self-administered questionnaires (SAQs) are surveys that are frequently used to gather information from participants. Many psychosocial constructs are assessed in this manner. These surveys can include those presented in paper/pencil format, administered through hand-held or laptop computer devices, or by cell phone. Audio computer-assisted self-interviews (ACASIs), computer-assisted self-interviews (CASIs), and computer-assisted phone interviews (CAPIs) are just a few examples of data collection methods that have been noted to increase the reliability and validity of information reported by participants (Morrison-Beedy, Carey, & Tu, 2006); these tools have become important options for researchers, especially when socially sensitive data, such as drug use, sexual behaviors, or criminal activity, need to be measured. Alternate methods for data collection can include interviews, conducted in person or by phone, where a research team member verbally presents the items to the participants and records their responses. Data collection may also involve the review of data already contained in written or computerized records; medical chart audits are a common example of this approach. System measures that assess constructs such as organizational change, quality, and conflict can be collected through electronic health system records, which are becoming more common in healthcare settings.

What can you do as an investigator if your participants are too young to tell you this information, if they cannot read or understand the language in which questions are written, if they are too ill or too frail to respond to questions, or if they are physically incapacitated and cannot write their responses? In these cases, data for your selected measures can be gathered using approaches that involve direct or indirect observation. Observational approaches can include filming and subsequent scoring of participant behaviors or activities. They also can include the use of checklists or other assessment measures completed by a research team member who is monitoring these behaviors or activities in-person, through a one-way mirror, or by review of audio or video tapes.

Physiological measures provide another option for measuring your study variables. Just about any aspect of human functioning can be quantified and measured although some approaches may be more costly than others. From the simple check of a participant’s blood pressure, pulse, or weight to the cardiac output of a patient running on a treadmill, to the intensity of a baby’s cry, assessing changes in physiological variables is a common measurement option. Measures that require the collection of biological specimens, such as cortisol, blood cells, genetic material, electrolytes, have become increasingly important ways to measure
the impact of interventions. These measures oftentimes provide a more objective measure than other approaches and may be new to some investigators. Thus, we provide additional details about biological measurement in Chapter 11. Still, it must be remembered that not all variables can or should be measured using physiological or biological methods. Many variables of interest (e.g., respect, self-worth, risk taking, intentions) cannot be assessed using these types of measures. You, as the investigator, must determine the most appropriate choice for collecting data on each variable in your study.

**TYPES OF SCALES**

A second level of measurement decision making involves the selection of response or answer choice format or rating scales. Some commonly used scales in intervention studies with self-report or observational measures include open-ended questions, Likert-type scales, diaries, visual analog scales (VAS), multiple answer options, and observational checklists. These are but a few of the vast choices of measurement found in the scientific literature; each has its strengths and limitations. Characteristics of the participants, data collection setting, and resources must be considered when selecting one option of measurement over the other. Figures 10.1 to 10.3 contain examples of some common scales.

Likert-type scales often assess the respondent’s agreement or belief in an item’s statement (e.g., “never,” “sometimes,” “always”). These scales can be used to quantify behavior (just not to the detail of interval-level scales). A VAS measurement is a horizontal line with anchors at each end. The study participant places a hatch mark on the line closest to the preferred choice in the range. It is important to determine the appropriate levels of measurement of every study variable. Researchers desire the highest level of measurement that can be obtained; however, it must be relevant for specific participants. For example, in determining educational status, participants can be measured using a variety of levels of measurement depending on the detail you wish to achieve (see Figure 10.1). Educational level can be assessed using a variety of levels of measurement. For instance, a dichotomous measure might simply pose: “completed high school” and “did not complete high school.” A categorical measure would state a wider range of choices such as “high school graduate,” “some college,” “college graduate.” More detailed data on educational level could be obtained using an interval-level measure. For example, participants could be asked to circle the highest grade they have completed from the following choices: “7-8-9-10-11-12.” Remember that it is easier to collapse data than to wish you had collected more detailed information during data collection time points.

Another important consideration in intervention research is to consider the order of administration of your questions. It is not uncommon for participants to have issues arise (e.g., feel ill, family member call, schedule conflict) that may prevent obtaining a completed survey. You may wish to avoid the more traditional order of “demographics, antecedents, and outcomes” and instead ask the participant to respond to “outcome” related items earlier in the battery of questionnaires rather than waiting until the end. This strategy helps to minimize the negative effect of participant fatigue. It also helps to increase the likelihood that you can collect outcome data even if the participant gets up and leaves. If participants decide they cannot complete the survey, then at least you have the core outcome data you seek to obtain.

Developing measures is a science unto itself and no easy task. Particularly for new investigators, selecting a measure that already exists with established validity and reliability enhances your chances of successfully designing, conducting, analyzing, and funding your intervention study. Try not to reinvent the wheel if at all possible—savings in time and effort should be considered. You also face the risk that your new measure will not be as useful or psychometrically strong as other current measures. Certainly, there are appropriate circumstances
THE NICU PARENTAL BELIEFS SCALE

Below are 18 statements that relate to you and your baby’s hospitalization. Hospital experiences differ for every parent. There are some parents who are not so sure about their baby’s needs and how they can best meet them while they are in the neonatal intensive care unit (NICU), while other parents are more sure about how to help their baby through this experience. There are no right or wrong answers to the following statements. Please circle the number that best describes your agreement or disagreement with each statement.

1. I know what characteristics and behaviors are common in premature babies hospitalized in the NICU.
   1 2 3 4 5
   Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

2. I am sure that what I do for my baby will be what is best to help him/her deal with being in the NICU.
   1 2 3 4 5
   Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

3. I feel comfortable in caring for my baby in the NICU.
   1 2 3 4 5
   Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

4. I know what characteristics and behaviors to expect in my baby while he/she is in the NICU.
   1 2 3 4 5
   Strongly disagree Disagree Neither agree or disagree Agree Strongly agree

FIGURE 10.1 Example of items from a Likert scale.


Think of the person you know best who has the AIDS virus. How similar are you and this person in the way you live. Make a mark anywhere along this line which answers this question the best.

---

Not at all like me.  Exactly like me.

FIGURE 10.2 An example of a visual analog scale.

when measures for the construct under investigation are not available or the state of the science indicates the need for such work. Such decisions always need to be weighed and carefully considered by the investigator.

**MEASUREMENT ERROR**

Measurement is not a perfect science, and all measures contain the possibility of error. The difference between the true value of a variable and the value obtained during data collection is the measurement error. Error can be classified as random, caused by chance, or systematic, a component of the measure itself. Researchers themselves may contribute to error in the measurement of a phenomenon. For example, participants may react to the researcher either positively (trying to please him or her during the course of the data collection) or negatively (not providing truthful or consistent responses because of fear or judgment). Common sources of error caused by the instrument itself include surveys with ambivalent or confusing questions, answer options that do not contain clearly delineated categories, or questions addressing sensitive issues that lead to response bias. Error in measurement can also occur because of environmental influences. Data collection that occurs in areas that are inconsistently lighted, noisy, or in temperature extremes can lead to error. Providing participants in the experimental intervention group with comfortable chairs and snacks while completing their assessments while placing your standard care control participants in a cold dark hallway for their data collection can contribute to measurement error. Finally, error can occur due to factors within the participants themselves—being in pain, fatigued, upset, or under the influence of substances—all contribute to subject error.

There are many ways in which to address these sources of error and increase the accuracy of measurement in your study. A fundamental approach is to use multiple measures...
for variables (e.g., self-report and observation), particularly those that target the primary outcome of interest. Data collection via blinded personnel who are not aware of which participants are assigned to which condition reduces the risk of differential bias. A critical component of conducting any research study, but often not on the mind or timeline of the investigator, is the need to standardize all study procedures and protocols, including standardization of measurement methods. Specific step-by-step instructions for all aspects of the study need to be documented. In particular, meticulous training of data collectors and assessing their competence prior to actual data collection in the field is imperative. This training can include observations by research team trainers or the principal investigator (PI), establishing inter- or intrarater reliability and mastery of equipment and mechanical devices. Strictly calibrated machines or instruments are essential as well as ongoing recalibration for standardization. In addition to using carefully trained data collectors, providing consistent environments for data collection is important. Meticulously editing, refining, and piloting instruments you may develop (or choose to use) to identify any items that are confusing, biased, incomplete, or too complex are critical to reducing measurement error.

**SUMMARY**

Selecting appropriate measures for an intervention study is the choice of the investigator. This choice can be both overwhelming (when there is a substantial array of measures for the variables of interest) and challenging (when the investigator must search extensively for even one appropriate measure). Developing a measure when one does not exist or when those that do are woefully inadequate is a science unto itself and requires intensive time and focus. Therefore, it is best to use already developed valid and reliable measures whenever possible. Remember, your intervention study will only be as good as the measures that assess study variables in a valid and reliable manner, so choose them carefully.

**Key Points From This Chapter**

- The investigator must decide which of the variables in his or her study is the primary outcome variable; power calculations and sample size are dependent on this choice.
- Levels of measurement (nominal, ordinal, interval, ratio) serve as the basis for type of statistical test used to evaluate intervention outcomes.
- A careful balance must be struck when it comes to selecting measures for your study—too few may leave you vulnerable to being unable to determine intervention effects; too many may increase participant burden as well as work for the research team.
- Specific, sensitive to change over time, appropriate, and feasible—these are all characteristics of useful measures in intervention work.
- Measurement error is both inherent and modifiable—careful attention must be paid to decreasing error related to researcher, instrumentation, environmental, and subject factors.
- Psychometrically sound measures are both reliable (consistently measuring constructs of interest) and valid (measuring the actual constructs of interest); evidence is needed to “support,” not to “prove” validity and reliability.
- Scientific instruments including self-report measures, biological, physiological, observational, and system assessments abound; the key to success is judicious, scientifically driven selection by the investigator.
REFERENCES


