Development and Pilot Testing of Four Desired Health Outcomes Scales

Laurel E. Radwin, RN, PhD, Michelle Washko, MS, Kathryn A. Suchy, RN, BS, and Kerri Tyman, RN, BA, BS

Purpose/Objectives: To develop and pilot test scales to measure desired health outcomes hypothesized to result from high-quality cancer nursing care: Fortitude Scale, Trust in Nurses Scale, Cancer Patient Optimism Scale, and Authentic Self-Representation Scale.

Design: Instrument development.

Setting: Community cancer support organization.

Sample: 66 recently treated patients with cancer who attended a cancer support organization workshop. The sample was predominately white, middle-aged, well-educated females.

Methods: Items for each scale were generated from qualitative data and the literature. The scales' properties were evaluated using expert panel assessment of content validity, cognitive interviews of patients with cancer, and reliability and validity testing of each scale with the Multitrait/ Multi-Item Analysis Program–Revised (MAP-R) statistical program.

Findings: Participant responses to the four scales did not include the lowest possible score. Responses yielded evidence of adequate Cronbach's alpha internal consistency reliability for each scale: 0.81 for the Fortitude Scale; 0.81 for Trust in Nurses Scale, 0.75 for Cancer Patient Optimism Scale, and 0.71 for Authentic Self-Representation Scale. The MAP-R statistics yielded evidence of acceptable convergent validity and discriminant validity.

Conclusions: The data provided preliminary evidence of acceptable psychometric properties for four scales designed to measure desired outcomes of cancer nursing care. Support was found for careful use of scales. Further psychometric testing with large samples is recommended.

Implications for Nursing: These scales represent an initial effort toward providing measures of the desired health outcomes that patients with cancer attributed to high-quality cancer nursing care.

Key Points . . .

➤ Current measures of the outcomes of nursing care mainly focus on adverse patient outcomes.

➤ Four scales were developed and pilot tested to measure desired health outcomes: Fortitude, Trust in Nurses, Cancer Patient Optimism, and Authentic Self-Representation scales.

➤ Preliminary evidence of acceptable psychometric properties was found for each scale.

Keeping patients safe is imperative, and nursing care affects patient safety (Page, 2004). Suboptimal nurse-staffing levels have been linked to adverse patient outcomes, including infections, decubitus ulcers, failure to rescue, and mortality (Aiken, Clarke, Sloane, Sochalski, & Silber, 2002; Mark, Salyer, & Wan, 2003; Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002). The goals of nursing care, however, extend beyond ensuring that care is safe and that adverse events do not occur. Goals of care include the achievement of desired health outcomes along with the avoidance of adverse outcomes (Committee on Quality of Health Care in America, Institute of Medicine, 2001). In the broadest sense, desired health outcomes are the positive intended results of care and include patients attaining optimal levels of physical functioning and intellectual performance, as well as psychosocial and emotional well-being (Harris-Wehling, 1990). The purpose of this study was to

develop and pilot test four scales to measure desired health outcomes that were hypothesized to result from high-quality cancer nursing care.

This study focused on the specific psychosocial and emotional outcomes identified in a middle-range theory of high-quality oncology nursing care. The theory was generated in a qualitative study with 22 patients with cancer (Radwin, 2000); theory concepts were designated as attributes or outcomes of high-quality oncology nursing care. Theory concepts generated as attributes included professional knowledge, continuity, attentiveness, coordination, partnership, individualization, rapport, and caring. The Oncology Patients' Perceptions of the Quality of Nursing Care Scale (OPPQNCS) subsequently was developed to measure the attributes described in the middle-range theory (Radwin, Alster, & Rubin, 2003).

The middle-range theory of high-quality oncology nursing care included five desired health outcome concepts: fortitude, a sense of well-being, trust, optimism, and authentic self-representation. The literature was searched for extant scales to measure these concepts. Berwick et al.'s (1991) five-item

Mental Health Inventory (MHI-5) provided an appropriate measure of the sense of well-being, which is defined in the middle-range theory as the patient's positive emotional state. However, an extensive literature review and correspondence with other researchers revealed that existing scales measured related, but not sufficiently similar, concepts for the remaining four desired health outcomes. Trust was defined in the middle-range theory as the patient's confidence that care was appropriate and reliable and would be as successful as possible. No scales to measure patients' trust in nurses were found. Although several scales measuring patients' trust in physicians were reviewed, items were specific to physicians', not nurses', role. For example, an item in Safran et al.'s (1998) Primary Care Assessment Scale asks whether respondents agree with the statement, "My doctor cares more about holding costs down than about doing what is needed for my health."

Fortitude was defined in the middle-range theory as the patient's strength and willingness to bear the effects of cancer treatments and the symptoms of the disease. Scales to measure the concepts of resilience (Wagnild & Young, 1993) and hardiness (Pollock & Duffy, 1990) were not sufficiently specific regarding the situational strength and stamina to undergo cancer treatments that defined fortitude in the middle-range theory. Optimism was defined as the patient's belief that he or she had made appropriate choices regarding treatment and the patient's feelings of hopefulness about treatment outcomes. Extant scales to measure hope (Herth, 1991; Miller & Powers, 1988) did not specifically address optimism that treatments would be successful. Authentic self-representation was defined as the patient's sense of genuine self-portrayal. Because one scale that measured authentic self-representation (Jack & Dill, 1992) did not address authenticity within a nurse-patient relationship, it could not be used. Therefore, to empirically test the middle-range theory of high-quality oncology nursing care, new scales were developed to measure the concepts of trust in nurses, fortitude, patient optimism, and authentic self-representation. These scales, combined with the OPPQNCS (Radwin et al., 2003) and the MHI-5 (Berwick et al., 1991), would allow for empirical testing of all of the concepts in the middle-range theory of high-quality oncology nursing care.

Methods

Four scales were devised and pilot tested in a convenience sample of 66 patients with cancer.

Generation of Items and Scales and Content Validity

Items were generated to measure each of the four desired health outcomes concepts described in the middle-range theory. Wording for the items was derived from data from the original qualitative study, existing scales that measured similar concepts, and a review of relevant literature. Items were developed as closed-ended declarative statements and were written in simple language geared toward an eighth-grade reading level. An eight-member expert panel reviewed the items for content validity and clarity and provided extensive comments. Items were revised accordingly, and new items were generated. These revised and new items were reviewed by an expert in survey research.

After approval by a university institutional review board, cognitive interviews were conducted with six patients with

cancer who recently received nursing care. The purpose of the interviews was to determine whether items were understandable, were answerable, and evoked the anticipated responses (Willis, 2002). Cognitive interviews involved reading each item aloud with a participant and asking probe questions such as "Would you please rephrase this question in your own words?" or "What situation were you thinking about when you answered this question?" Participants' responses were reviewed with two methods experts. Based on the cognitive interviewing data, some scale items were eliminated and new items were generated.

The resultant scales for pilot psychometric testing contained the following number of items: fortitude, four items; trust in nurses, six items; patient optimism, four items; and authentic self-representation, four items. A general question was included in each scale (i.e., "Please rate how much you trusted your nurses on a scale from 1–10, with 1 indicating that you did not trust your nurses at all and 10 indicating that you trusted your nurses as much as possible.").

Each item represented a nurse activity, patient activity, or patient feeling. Respondents ranked the frequency of the activity or feeling on a Likert scale (1 = never, 2 = rarely, 3 = some of the time, 4 = a good bit of the time, 5 = usually, 6 = always). The following items were reverse scored: "How often did you think that your nurses did not tell you information that you would have liked to have known?" "When with your nurses, how often did you act as if you felt better than you really felt?" "How often have you felt that you could not complete all of the cancer treatments that you wanted to complete?" "How often have you felt that you were not strong enough to get through what you were facing?" "How often have you felt that your situation will not turn out well?" and "How often have you felt pessimistic about your future?"

Psychometric Testing of the Scales

Data collection: Attendees at workshops and a health fair offered by a New England community-based cancer support organization during six months in 2003 were invited to participate in the study. Potential participants were aged 18 or older, were able to read English, indicated that they had received nursing care in a hospital or clinic in the past six months, and were sufficiently healthy to attend an educational, fitness, or networking workshop or a health fair. A research team member explained the study as follows.

We are planning to conduct a study to examine the relation between the quality of cancer nursing care and outcomes experienced by people who receive nursing care. Prior to conducting the study, we are developing questionnaires to measure some of the outcomes. We are inviting persons with cancer who have been cared for by nurses in the last six months to participate. If you agree to participate in this study about outcomes, you will be asked to fill out a few questionnaires, which will take about 20 minutes of your time. If you are interested, please take a packet as you leave this program. If you have questions or need help completing these forms, please let us know. We are pleased to provide a $5 appreciation fee to the [community-based cancer support organization] for each person who completes the questionnaires.

Interested potential participants were given packets containing the outcome scales, a demographic questionnaire, a pencil,

and an informational letter about the study. Study participants' informed consent was assumed with completion and return of the packet.

Data analysis: The psychometric properties of each scale were assessed using Multitrait/Multi-Item Analysis Program–Revised for Windows® software (Ware, Harris, Gandek, Rogers, & Reece, 1997). Responses were converted to a 0–100 scale so that easily comparable descriptive statistics could be calculated for the items and scales. Each scale was examined for internal consistency reliability using Cronbach's alpha, and item and scale descriptives were calculated, including floor and ceiling effects. Assumptions of Likert scaling, such as item-convergent validity and item-discriminant validity, also were assessed.

Findings

Sample

Sixty-six of the 123 questionnaires that were distributed were returned, yielding a response rate of 54%. Approximately 77% of the participants were female. Participants ranged in age from 25–80 years ($\overline{X}$ = 53.3, SD = 12.57). Most participants (97%) were white, 3% were Asian, and 1.5% were Hispanic or Latino. Most participants (80%) had completed college. About 25% of the participants had an annual total household income of as much as $39,999; 15% reported $40,000–$60,000; and 53% reported more than $60,000 (7.6% did not respond). When asked about the cancer's origination, 49% of the participants designated the breast, 12% designated other, 9% selected the bone marrow, and 8% each reported the lungs and lymph nodes. About 75% received chemotherapy, 62% had cancer surgery, 58% had radiation therapy, and 15% had immunotherapy.

Item-descriptive statistics (i.e., missing data, response frequencies, means, standard deviations, and range) were calculated to identify items that were difficult to answer (missing data) and items that did not reflect expected variability. All values were in range and reflected variability. Internal consistency reliability for each scale was calculated using Cronbach's alpha, and a criterion of 0.70 was set (Nunnally & Bernstein, 1994). Cronbach's alpha was 0.81 for the Fortitude Scale and 0.75 for the Cancer Patient Optimism Scale. Alpha-if-item-removed statistics indicated that no item lowered the reliability statistic in either scale. Alpha-if-item-removed statistics for the Trust in Nurses Scale indicated that the removal of one item (i.e., "How often did you think that your nurses did not tell you information that you would have liked to have known?") resulted in an increase in the Cronbach's alpha from 0.76 to 0.81. Similarly, removal of one item (i.e., "When with your nurses, how often did you act as if you felt better than you really felt?") increased the Cronbach's alpha for the Authentic Self-Representation Scale from 0.53 to 0.71. These two items were removed, and analysis proceeded using the remaining 16 items.

Item descriptives for the 16 items are displayed in Table 1. Missing data ranged from 0–3 respondents (0%–4.5%), suggesting that items, for the most part, were not difficult to answer or were not objectionable. The possible range for all items and scales is 0–100. In this pilot sample, item means ranged from 57.50 (SD = 29.28) for "How often did you tell your nurses how you were truly feeling about the cancer?" (from the Authentic Self-Representation Scale) to 91.52 (SD = 14.06) for "How often did you believe that your nurses were acting in your best interest?" (from the Trust in Nurses Scale). Scale means ranged

from 67.62 (SD = 20.42) for the Authentic Self-Representation Scale to 87.75 (SD = 12.36) for the Trust in Nurses Scale. Ceiling and floor effects (i.e., percentage of responses at the highest response category and percentage of responses at the lowest response category, respectively) demonstrated that none of the scales had a complete range of scores from 0–100 because each lacked the lowest possible score. The lowest scores on the Fortitude, Trust in Nurses, Cancer Patient Optimism, and Authentic Self-Representation scales were 35, 48, 35, and 26.67, repectively. Respondents at the ceiling (100) were 11.1% for the Fortitude Scale, 19.0% for the Trust in Nurses Scale, 1.6% for the Cancer Patient Optimism Scale, and 6.3% for the Authentic Self-Representation Scale.

Likert-scaling assumptions were evaluated using specific procedures recommended by Ware et al. (1997). Three assumptions relate to the validity of the items and scales: item-convergent validity, item-discriminant validity, and assumptions regarding internal consistency intrascale correlations. Convergent validity reflects whether different measures of the same concept correlate with one another. Each item in a Likert scale is a single measure of the concept operationalized by the entire scale, and convergent validity can be evaluated by the item-scale correlation. In this study, the criterion for item-scale correlation of 0.40 was set to denote satisfactory item-scale convergent validity (Hays & Hayashi, 1990). Calculations included a correction for overlap so that inflated estimates of item-scale correlations were not made. All of the items had item-scale correlations of more than 0.40 except "How often were you comfortable being your true self with your nurses?" This was found in the Authentic Self-Representation Scale and had a 0.30 correlation.

Discriminant validity reflects whether measures for different concepts differ from one another. Item-discriminant validity is supported when an item correlates most strongly with the group of items hypothesized to measure the same concept and, simultaneously, the item correlates less strongly with items hypothesized to measure other concepts (Hays & Hayashi, 1990). The standard error for the correlation sets the criterion. An item that scores –2 standard errors of the correlation is considered a scaling failure because that item is statistically and significantly correlated less strongly with items hypothesized to be in the same scale and correlated more strongly with items hypothesized to be in different scales. In this study, items in each of the four scales scored more than –2 standard errors, indicating acceptable discriminant validity for each scale.

Evidence of measurement of a distinctive concept is provided when the internal consistency of the scale as measured by Cronbach's alpha is higher than the correlation of that scale with the other scales. As seen in Table 2, the Cronbach's alpha for each scale (on the diagonal) considerably exceeded the correlation of that scale with the other scales (off diagonal). Thus, support was found that each scale measured a distinct concept.

The practice of simple summing of items in a Likert scale is based on the assumption that each item measures an essentially equivalent amount of the underlying concept. This equivalence is reflected in relatively similar item-scale correlations for items in a scale. As seen in Table 1, similar item-scale correlations were found for items in the same scale. The notable exception was "How often were you comfortable being your true self with your nurses?" in the Authentic Self-Representation Scale.

Table 1. Descriptive Statistics and Correlations for the Outcomes Scales

Item	X̄	SD	Range	Missing (%)	Item-Scale Correlation	Floor (%)	Ceiling (%)
Fortitude Scale	81.48	14.52	35–100	–	–	0.0	11.1
• How often have you felt that you could not complete all of the cancer treatments that you wanted to complete?	87.62	20.46	0–100	4.5	0.69	–	–
• How often have you felt that you were not strong enough to get through what you were facing?	76.25	18.13	20–100	3.0	0.73	–	–
• How often have you felt strong enough to handle the difficulties that the cancer has caused?	76.25	17.77	0–100	3.0	0.68	–	–
• How often have you had enough emotional energy to continue with your cancer treatment?	86.46	15.85	40–100	1.5	0.42	–	–
Trust in Nurses Scale	87.75	12.36	48–100	–	–	0.0	19.0
• How often were your nurses there when you needed them?	91.39	15.80	40–100	1.5	0.59	–	–
• How often did you believe that your nurses were acting in your best interest?	91.52	14.06	40–100	0.0	0.73	–	–
• How often did you trust what your nurses told you?	84.62	15.72	40–100	1.5	0.57	–	–
• How often did your nurses do what they said they would do?	88.00	13.13	40–100	1.5	0.59	–	–
• How often did your nurses provide accurate information about the cancer?	83.49	21.41	0–100	4.5	0.61	–	–
Cancer Patient Optimism Scale	72.62	13.01	35–100	–	–	0.0	1.6
• How often have you felt that your situation will not turn out well?	68.62	16.57	20–100	1.5	0.65	–	–
• How often have you felt that you made the right cancer treatment choices?	84.24	15.50	20–100	0.0	0.44	–	–
• How often have you felt that you would get the results that you wanted from your cancer treatment?	72.42	19.14	0–100	0.0	0.50	–	–
• How often have you felt pessimistic about your future?	65.94	17.00	20–100	3.0	0.59	–	–
Authentic Self-Representation Scale	67.62	20.42	26.67–100	–	–	0.0	6.3
• How often did you ask your nurses questions about the cancer that were truly on your mind?	65.31	27.66	0–100	3.0	0.62	–	–
• How often did you tell your nurses how you were truly feeling about the cancer?	57.50	29.28	0–100	3.0	0.76	–	–
• How often were you comfortable being your true self with your nurses?	79.69	17.85	40–100	1.5	0.30	–	–

Discussion

Pilot testing of the four scales provided evidence of acceptable preliminary internal consistency reliability, item-convergent validity, and item-discriminant validity. Intrascale correlations compared with scale Cronbach's alphas indicated that each scale measured a distinctive concept. Performance of the Authentic Self-Representation Scale could have been improved with the removal of one item. The authors determined that this item should be retained because of its conceptual importance to the scale and the belief that performance of each scale should be assessed in larger samples before additional items are eliminated.

Table 2. Reliability Coefficients and Intrascale Correlations

Scale	Fortitude	Trust in Nurses	Cancer Patient Optimism	Authentic Self-Representation
Fortitude	(0.81)	–	–	–
Trust in Nurses	0.20	(0.81)	–	–
Cancer Patient Optimism	0.43	0.33	(0.75)	–
Authentic Self-Representation	0.27	0.45	0.20	(0.71)

The data indicated that no respondents scored at the floor of each scale. Ware et al. (1997) noted that, at times, the range of scores on a scale reflects the group being studied rather than a scale property. For example, in this study, the finding that no respondents scored lower than 35 on the Cancer Patient Optimism Scale may indicate that some degree of optimism may be felt by patients with cancer, no matter how discouraged or depleted they may be.

In addition, the mean for the Trust in Nurses Scale was particularly high (87.75). Gallop poll results repeatedly have indicated that nurses, as well as several other professions, fare well with the American public when ranked for honesty and ethics (Gallup Organization, 2003). The high mean on the Trust in Nurses Scale may indicate that participants in this study had experiences with nurses whose honesty and ethics inspired trust.

Limitations

Data describing nonparticipants were not collected because demographic information about all of the workshop and health fair participants was not available. Therefore, a potential source of bias included that nonparticipants may have differed significantly from the participants in race, ethnicity, age, education, income, type of cancer, and treatment.

Implications

In *Crossing the Quality Chasm: A New Health System for the 21st Century* (Committee on Quality of Health Care in America, Institute of Medicine, 2001), a select panel from the Institute of Medicine proclaimed that patients do not receive the care they need, want, and deserve. The panel called for a renewed commitment to high-quality health care—care that enhances the achievement of desired health outcomes. Nevertheless, measure sets devised to evaluate the quality of nursing care remain heavily weighted toward the measure of adverse outcomes. For example, the patient-centered nursing outcome measures proposed by the National Quality Forum (2004) included failure to rescue, pressure ulcer prevalence, hospital- and ventilator-acquired pneumonias, use of restraints, urinary tract and urinary catheter-associated infections, and central line catheter infections. Moreover, although the Oncology Nursing Society Nursing Sensitive Outcomes Project Team (Given et al., 2003) included functional measures in its recommended outcomes data, the majority of outcomes related to a decrease in physical symptoms and psychological distress, rather than an increase in desired outcomes such as optimism, fortitude, and authentic self-representation. Measures of desired outcomes represented by the four scales in this pilot study, as well as a measure of a sense of well-being, supplement current efforts to evaluate oncology nursing care by offering opportunities to monitor the achievement of positive patient states as well as the reduction or elimination of adverse patient states.

The scales proposed here represent a first step toward measuring the desired health outcomes that patients with cancer themselves ascribe to high-quality nursing care. Further psychometric testing of the scales is needed; however, these scales exhibit sufficient preliminary psychometric properties to warrant their careful use. Plans call for use of these scales in a study with a large sample funded by the Agency for Health Care Research and Quality (Radwin, 2001). Psychometric properties of the scales will be examined further in this study, which is one effort to examine nursing's impact on a full range of desired health outcomes and provide evidence on nursing contributions to high-quality cancer care.

Author Contact: Laurel E. Radwin, RN, PhD, can be reached at laurel .radwin@umb.edu, with copy to editor at rose_mary@earthlink.net.

References

Aiken, L.H., Clarke, S.P., Sloane, D.M., Sochalski, J., & Silber, J.H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA, 288,* 1987–1993.

Berwick, D.M., Murphy, J.M., Goldman, P.A., Ware, J.E., Jr., Barsky, A.J., & Weinstein, M.C. (1991). Performance of a five-item mental health screening test. *Medical Care, 29,* 169–176.

Committee on Quality of Health Care in America, Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: National Academy Press.

Gallup Organization. (2003). Public rates nursing as most honest and ethical profession. Retrieved December 19, 2003, from http://www.gallup.com/subscription/?m=f&c_id=14141&Version=p

Given, B., Beck, S., Etland, C., Gobel, B.H., Lamkin, L., & Marsee, V.D. (2003). Oncology Nursing Society nursing sensitive outcomes. Retrieved December 7, 2004, from http://www.ons.org/research/outcomes/pdf/draft.pdf

Harris-Wehling, J. (1990). Defining quality of care. In K. Lohr (Ed.), *Medicare: A strategy for quality assurance: Volume 2: Sources and methods* (pp. 116–139). Washington, DC: National Academy Press.

Hays, R.D., & Hayashi, T. (1990). Beyond internal consistency reliability: Rationale and user's guide for Multitrait Scaling Analysis Program on the microcomputer. *Behavior Research Methods, Instruments, and Computers, 22,* 167–175.

Herth, K. (1991). Development and refinement of an instrument to measure hope. *Scholarly Inquiry for Nursing Practice, 5,* 39–51.

Jack, D.C., & Dill, D. (1992). The Silencing the Self Scale: Schemas of intimacy associated with depression in women. *Psychology of Women Quarterly, 16,* 97–106.

Mark, B.A., Salyer, J., & Wan, T.T. (2003). Professional nursing practice: Impact on organizational and patient outcomes. *Journal of Nursing Administration, 33,* 224–234.

Miller, J.F., & Powers, M.J. (1988). Development of an instrument to measure hope. *Nursing Research, 37,* 6–10.

National Quality Forum. (2004). National Quality Forum endorses national voluntary consensus standards for nursing-sensitive performance measures and endorses two additional nursing home performance measures. Retrieved February 15, 2004, from http://www.qualityforum.org/pr nursingcarevcsFINAL1-30-04.pdf

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine, 346,* 1715–1722.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Page, A. (Ed.). (2004). *Keeping patients safe: Transforming the work environment of nurses.* Retrieved February 14, 2004, from http://books.nap.edu/catalog/10851.html

Pollock, S.E., & Duffy, M.E. (1990). The health-related hardiness scale: Development and psychometric analysis. *Nursing Research, 39,* 218–222.

Radwin, L. (2000). Oncology patients' perceptions of quality nursing care. *Research in Nursing and Health, 23,* 179–190.

Radwin, L. (2001). *Testing the Quality Health Outcomes Model in cancer care* [Grant K08 HS11625 proposal to United States Public Health Service, Agency for Health Care Research and Quality].

Radwin, L., Alster, K., & Rubin, K.M. (2003). Development and testing of the Oncology Patients' Perceptions of the Quality of Nursing Care Scale. *Oncology Nursing Forum, 30,* 283–290.

Safran, D.G., Kosinski, M., Tarlov, A.R., Rogers, W.H., Taira, D.H., Lieberman, N., et al. (1998). The Primary Care Assessment Survey: Tests of data quality and measurement performance. *Medical Care, 36,* 728–739.

Wagnild, G.M., & Young, H.M. (1993). Development and psychometric evaluation of the Resilience Scale. *Journal of Nursing Measurement, 1,* 165–178.

Ware, J.E., Jr., Harris, W.J., Gandek, B., Rogers, B.W., & Reece, P.R. (1997). *MAP-R for Windows: Multitrait/Multi-Item Analysis Program.* Boston: Health Assessment Lab.

Willis, G. (2002). *Cognitive interviewing and questionnaire design.* St. Petersburg Beach, FL: American Association for Public Opinion Research.