

A Comprehensive Strategy for Developing Closed-Ended Survey Items for Use in Studies of Older Adults

Neal Krause

School of Public Health and Institute of Gerontology, University of Michigan, Ann Arbor.

Objective. The purpose of this study is to describe the development and implementation of a nine-step strategy for devising closed-ended survey questions that assess religion in late life. The intent was to illustrate one way in which qualitative and quantitative methods could be combined in the same study.

Methods. The following methods and procedures were developed to create closed-ended questions: Focus groups, in-depth interviews, input from ongoing quantitative studies, input from an expert panel, cognitive interviews, a quantitative pretest, a nationwide random probability sample of elderly people, and rigorous empirical psychometric testing. Three hundred ninety-nine older people took part in the first seven steps, and 1,500 elders participated in the nationwide survey.

Results. Approximately 175 closed-ended survey items were developed assessing 14 different major dimensions of religion. In the process, practical solutions to a number of problems encountered in implementing the nine-step strategy are discussed.

Discussion. The item development strategy may serve as a template that can be used to improve the quality of closed-ended survey items that assess a wide range of topics in social gerontology.

A growing number of investigators are calling for studies that combine qualitative and quantitative research methods (Morgan, 1998). This promising strategy is consistent with the principle of triangulation, in which the combined strengths of two or more methods are used to produce more valid results than would be obtained by using the same methods in isolation (Denzin, 1970). However, as Morgan (1998) points out, there are a number of ways to combine qualitative and quantitative methods in the same study. In fact, he proposes four broad approaches for configuring the interface between these methodologies. One is especially important for the purposes of the present study. In particular, Morgan (1998) argues that the insights provided by qualitative methods, such as focus groups, may be useful for crafting high-quality, closed-ended survey questions.

Unfortunately, researchers often encounter two problems when they try to use both qualitative and quantitative methods for this purpose. First, even though many investigators call for the use of qualitative methods to develop quantitative survey items, relatively few show precisely how this may be accomplished. Instead, most researchers merely state they have used some qualitative procedure, such as focus groups, to draft closed-ended survey items, but they do not fully describe the steps that were followed (see Fultz & Herzog, 1993, for a notable exception). Second, a number of different qualitative methods are available, including focus groups, in-depth interviews, and participant observation studies. Each of these procedures has its own unique strengths and weaknesses. Consistent with the principle of triangulation, it would, therefore, appear that a well-developed item development strategy should include more than one qualitative

method. However, no one in social gerontology has discussed how this should be done.

The purpose of the present study is to address these gaps in the literature by providing a comprehensive strategy for combining multiple qualitative methods to develop closed-ended survey items. This strategy begins with focus groups and culminates in the quantitative analysis of closed-ended survey items that were administered to a nationwide sample of older people. Throughout, an emphasis is placed on providing practical, hands-on advice for bridging the two methodological approaches.

Three points should be kept in mind as this item development strategy is reviewed. First, qualitative methods are used in this study to enhance the effectiveness of quantitative survey questions. This may create the mistaken impression that qualitative methods occupy a secondary position in the research enterprise, and that their sole purpose is to provide raw grist for quantitative work. Nothing could be further from the truth. Instead, it should be emphasized that qualitative methods are important in their own right. In fact, Morgan (1998) discusses how quantitative methods can be used to enrich findings from studies that have a primary qualitative focus. Second, the research described later was designed to develop survey items to assess religion in late life. Even so, these procedures can be readily adapted to virtually any substantive domain. Finally, a number of the methodologies that follow have already appeared in the literature (e.g., focus groups and in-depth interviews). However, no one has pulled them together to form a single comprehensive strategy for developing better closed-ended survey items.

The discussion that follows is divided into four main

sections. First, the benefits of combining qualitative and quantitative methods to develop closed-ended survey items are explored in greater detail. Second, as noted previously, the substantive goal of the present study was to develop a comprehensive battery of items to measure religion in late life. To provide a better context for understanding how these measures were devised, problems and issues in the assessment of religion are discussed briefly. Methodological issues that cut across all nine steps in the proposed item development strategy are discussed at this juncture as well. Third, the nine-step item development strategy is presented in detail. Fourth, limitations and unresolved issues in the implementation of the proposed strategy are discussed.

Benefits of Combining Qualitative and Quantitative Methods

The advantages of using qualitative and quantitative methods in the same study are best presented by briefly reviewing how researchers typically select or develop closed-ended survey items. One frequently used approach is to simply take scales or items that are available in the literature. Although this is certainly a reasonable strategy, it is often done without checking to see if the items were developed specifically for older people. Research on social desirability provides a good example of what may happen under these circumstances.

Survey researchers often ask questions about sensitive issues, such as the excessive use of alcohol or sexual difficulties. Some investigators are concerned that instead of answering these questions honestly, study participants may respond in a socially appropriate manner (DeMaio, 1984). In an effort to deal with this problem, scales have been devised to identify those who are likely to give socially desirable responses. Perhaps the most widely used measure is the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964). This measure has been used in a number of studies with older adults (see, e.g., Kozma & Stones, 1987), yet a careful review of the history of this scale reveals that it was developed primarily with undergraduate students. As a result, this measure contains a number of items that do not seem appropriate for elderly populations (e.g., "If I could get into a movie without paying and be sure I was not seen, I would probably do it"; Crowne & Marlowe, 1964, p. 23).

Another common strategy for developing quantitative questionnaires involves writing survey items from scratch with input from colleagues working in the field. A more sophisticated variant of this approach involves convening a panel of experts (Fetzer Institute & the National Institute on Aging, 1999). Here, the group of experts is charged with identifying the key dimensions of a conceptual domain and devising good survey measures of them. Unfortunately, researchers who rely on expert panels rarely take steps to verify that all key dimensions of a conceptual domain have been identified, or that the dimensions that have been specified by the experts are valid. Instead, the basis of this strategy rests largely on consensus. The problem with relying on consensus is captured succinctly in Faust and Minor's (1986) critique of the *DSM-III* (*Diagnostic and Statistical Manual of Mental Disorders*, ed. III; American Psychiatric Association, 1980). The purpose of the *DSM-III* was to devise a comprehensive scheme to identify, categorize, and describe the different types

of clinical mental disorder. This was accomplished, in part, by turning to panels of experts. But as Faust and Minor (1986) point out, expert consensus is not necessarily valid: there was a time in history when most experts believed the world was flat.

Despite the criticisms of Faust and Minor (1986), care must be taken not to disparage the use of expert panels. In fact, this valuable procedure is incorporated in the item development strategy discussed later in this article. Instead, the key issue is that problems may arise when expert panels become the sole means of developing closed-ended survey items.

Proponents of the two item development strategies discussed previously might argue that problems in their work may be detected and resolved through careful pretesting. This may be true, but pilot tests are typically not conducted in a rigorous way. Instead, a draft of the closed-ended questionnaire is typically administered to between 25 and 75 respondents (Oksenberg, Cannell, & Kalton, 1991). The questionnaire is administered by interviewers who are instructed to look out for two types of problems. First, they are told to flag any difficulties study participants have in answering the questions. Second, the interviewers are instructed to report any problems they encounter in administering the interview schedule (e.g., problems following skip-patterns). Once the pilot interviews have been conducted, a debriefing session is held by a field manager, who systematically works through each question in an effort to see if any of the survey items are not working well. Unfortunately, as Oksenberg and her colleagues point out, this strategy is not entirely satisfactory because it relies on a respondent's ability to express a problem and the interviewer's ability to detect that one is present (Oksenberg et al., 1991). Moreover, it is not clear what to do when interviewers disagree in debriefing sessions about the presence or the nature of a problem. Finally, standard pretesting procedures may flag items that are in need of revision, but this strategy often provides little information about how to fix these problems.

It is precisely for these reasons that qualitative studies may be especially useful for developing closed-ended survey items. By allowing people to talk freely without imposing a researcher's prior assumptions on the study, qualitative methods provide an excellent opportunity for getting direct access to a substantive domain, such as religion. Listening while older adults share their experiences and feelings about religion gives researchers an improved understanding of this construct from the participants' perspective. As a result, investigators are in a better position to uncover broad themes and content areas that more accurately reflect the ways that elders themselves think about and practice religion in daily life, thereby making it possible to identify key dimensions of a phenomenon that have not appeared previously in the literature.

Another major advantage of qualitative methods is that they allow investigators to capture the words and phrases that older adults use when talking about things like religion. These words and phrases provide excellent raw grist for writing closed-ended question stems. As a result, researchers may enhance the relevance and comprehension of the quantitative survey measures they devise, thereby improving the validity of study findings.

Issues in the Measurement of Religion in Late Life

A compelling number of studies indicate that older adults who are religious may enjoy greater health and well-being than elderly people who are less involved in religion (see Koenig, McCullough, & Larson, 2001, for a comprehensive review of this literature). Unfortunately, there are at least two problems with this research. First, as Ellison and Levin (1998) point out, many investigators use relatively crude measures to assess religion. In particular, religious preference or the frequency of church attendance are often used as the sole indicators of religion. Although some investigators have examined other facets of religion, such as religious support (Krause, Ellison, & Wulff, 1998) or religious coping (Pargament, 1997), relatively few efforts have been made to systematically stake out the entire content domain of this elusive construct (for a notable exception, see the work of the Fetzer Institute & the National Institute on Aging, 1999).

The second problem with research on religion and health in late life arises from the fact that the wide majority of religion measures have been developed solely with younger people. This is problematic because a number of researchers argue that the nature and meaning of religion may change as people grow older (Koenig, 1994). But even when measures have been developed specifically for older adults, they may not be appropriate for all elderly people. More specifically, there is growing evidence of significant differences in the way that older White persons and elderly Black persons view and practice religion (Chatters, 2000). Yet, there do not appear to be any studies that have systematically developed measures of religion that are appropriate for older people in both racial or ethnic groups.

The present study was designed to meet these problems head-on by developing a comprehensive battery of items to measure religion in late life that would be appropriate for use with both Black and White respondents. In the process of conducting this research, a number of critical decisions had to be made to make the scope of this task more manageable and to ensure that the data adequately capture the views of the typically older person. This was accomplished by carefully specifying the exclusion criteria and developing a sound plan for sampling study participants.

Three exclusion criteria were used in this research. First, the study was restricted to individuals who were currently practicing Christians, were Christians in the past but no longer practice any religion, or people who were not involved with any faith at any point in their lifetime. However, since 90% of older adults in the United States currently affiliate with the Christian faith, the measures should be useful for studying the wide majority of elderly people (Princeton Religion Research Center, 1994). Second, potential study participants were screened to see if they were either members of the clergy or if they resided with someone who was a member of the clergy. Those who were ministers or people who lived with a member of the clergy were excluded from the study for the following reasons. Views on the meaning, nature, and practice of religion among clergy are likely to differ from those of the average lay person because of the extensive formal training pastors receive in seminary school. In addition, individuals who live with a member of the clergy were excluded because they often play a role in the church

that differs significantly from that of the typical rank-and-file church member (e.g., the minister's wife). Finally, the third exclusion criterion had to do with religious involvement. In particular, potential study subjects were asked the following question: "How important is religious faith in your life? Would you say it is very important, somewhat important, a little important, or not at all important?" Individuals were excluded from the initial steps of the study if they indicated that religion was not at all important to them. This was done because it did not make sense to ask people who are not at all religious to discuss how they practice religion in their daily lives. Excluding those who are not religious is not likely to be a major problem because research indicates that only about 8% of older adults say that religion is not very important to them (Gallup, 1999). Even so, it should be emphasized that once a preliminary set of closed-ended items was developed, all study measures were carefully tested with all older subjects, regardless of how religious they may be.

Once the exclusion criteria were specified, a decision had to be made about how to select eligible participants for the qualitative components of the study. There is a fairly large literature on qualitative sampling methods (see, e.g., Luborsky & Rubinstein, 1995), but this is not the place to discuss the relative strengths of qualitative and quantitative sampling procedures. Instead, it makes more sense at this point to review the reasons why quantitative, random-probability sampling procedures were used in this study. Even though this study was restricted to older Christians, there is tremendous diversity within the Christian faith. In addition to the obvious differences between Protestants and Catholics, there is considerable variation in the way that Protestants practice their faith. The fact that the General Social Survey codes religious preference into more than 20 different Protestant churches or denominations provides ample evidence of this (see the following website for a detailed overview of the General Social Survey: <http://www.icpsr.umich.edu/gss/>). Initially, it might appear that selecting potential respondents from different Protestant churches would be a good sampling strategy. However, this is not a viable option because findings from the nationwide survey conducted for the present study reveal that approximately 24% of eligible study participants either do not go to church at all, or attend church only once or twice a year. Subsequent analyses revealed that some were unable to do so because they were either too ill to go to church, or because they were caring for someone who was sick.

Given these constraints, a decision was made to use random-probability sampling procedures. The study population was defined as all individuals who are noninstitutionalized, English-speaking, and at least 65 years of age and older. Geographically, the study population was restricted to all eligible persons who reside in Washtenaw County, MI. The sampling frame consisted of all eligible individuals who were contained in the Health Care Financing Administration (HCFA) Medicare Beneficiary Eligibility List (HCFA is now called the Centers for Medicare and Medicaid Services). This list contains the name, address, sex, and race of virtually every older person in Washtenaw County. It should be emphasized that people are included in this list even though they are not currently receiving Social Security benefits.

However, some older adults are not covered in this database because they do not have a Social Security number (this may be caused by factors such as illegal immigration).

The names of potential study participants were selected from the HCFA list with a simple random sampling procedure. Then, sampled individuals were sent a letter informing them about the nature of the study, and indicating that a member of the research team would be calling them shortly. The purpose of the telephone call was to briefly screen sampled individuals to see if they met the eligibility criteria discussed previously. In general, phone numbers for study participants were found by checking the local telephone directories. However, some telephone numbers were more difficult to locate. Although precise counts were not kept, it seemed that telephone numbers were more difficult to obtain for older Black persons than older White persons. This problem was handled by turning to the recent addition of the *R. L. Polk City Directory* for the catchment area covered by our study. This book contains the results of a survey conducted by the Polk Corporation to obtain information (including phone numbers) for every individual residing in a given geographical area.

METHODS

The nine-step strategy for developing closed-ended survey items that was developed for this study is presented in Table 1. The steps in this table are presented in the order in which they were executed. In the discussion provided, the procedures that were followed in implementing each step are described in detail. In addition, where it is appropriate, substantive findings are presented to illustrate precisely how the procedures were implemented.

Focus Groups

All together, 8 focus groups were conducted with a total of 63 older adults. Thirty-one were older African Americans and 32 were elderly White adults. The groups were run in four pairs—one consisting solely of older White adults and one made up entirely of elderly Blacks adults. New subjects were recruited for each round of focus groups. Focus group moderators were matched to the race of the study participants. All subjects were paid \$25 for participating in this study (subject remuneration was used in all phases of data collection in this study). David Morgan, an expert on focus groups (Morgan & Krueger, 1998), came to Ann Arbor to train the moderators and to help run the first round of focus groups.

The focus groups were conducted in a local hotel. All focus groups were tape recorded and transcribed. In reviewing

focus group transcripts, it is often useful to know the identity of the speaker because this information enables researchers to get a better sense of important issues like how widely opinions are shared in the group. Unfortunately, when listening to tapes of a focus group that consists of six or eight people, it is often difficult to know who is speaking. To attribute responses to specific people, a court stenographer was hired to transcribe all focus group sessions. Because court stenographers are trained to produce virtually flawless transcripts, the risk of encountering transcription errors was reduced significantly. Initially, it may appear that study subjects would be bothered or inhibited by the presence of a stenographer in the focus group sessions. There are two reasons why this was not a problem in the present study. First, field notes taken during the course of the focus groups revealed that study participants paid little attention to the stenographer and rarely even looked at this individual during the sessions. Second, at the end of the focus groups, study participants were asked if they were bothered in any way by the presence of the court recorder. Without exception, they indicated this was not the case, and some even made jokes about it.

Two key concepts guided the flow of focus group discussions: The funnel approach (Morgan, 1988) and the saturation point (Glaser & Strauss, 1967). Because the focus groups were conducted sequentially over time, a funnel approach was used to devise the moderator's guide. For those unfamiliar with focus groups, the moderator's guide contains a list of topics or questions that are used to stimulate focus group discussions. The first round of focus groups began with very general questions that were designed to throw as broad a net as possible, thereby ensuring that the views of the research team were not imposed on the group discussions (e.g., "What is the most important part of living a religious life?"). Initially, the members of the research team read the transcripts from the first round of focus groups on their own. Then, the team met as a group to reach a consensus about what had been said. This information was subsequently used to draft a series of more specific questions for the next round of focus groups. So, for example, more targeted questions were asked about church-based social support ("Some people say that the help and guidance they get from people at church is important. What do you think? What are some of the ways people in your church may help each other?").

Two important points must be made about the funnel approach. First, consistent with the basic tenets of qualitative interviewing (Madill, Jordan, & Shirley, 2000), this strategy helps ensure that the substantive content of the focus group questions, and the way they are phrased, are determined by the subjects. Second, because data from earlier rounds of focus groups were used to devise questions for later rounds of focus groups, the content of the moderator's guide changed several times during this phase of the study (see May 1991, for a more detailed discussion of this data collection strategy).

Concerns may arise over the use of the funnel approach because it may seem as though the use of more focused questions would prohibit those who participate in later rounds of focus groups from expressing their own views. This potential problem was addressed in the following manner. Even in

Table 1. Steps for Developing Closed-Ended Survey Questions

1. Focus groups
2. In-depth interviews
3. Input from quantitative studies
4. Developing preliminary quantitative measures
5. Review by expert panel
6. Cognitive interviews
7. Pilot study
8. Nationwide survey
9. Psychometric testing

later rounds of focus groups, the moderator always began with general questions about religion, and only asked more focused questions after study participants had sufficient time to respond to the more general questions. In this way, the general questions provided a way to continually bring new information on religion to the foreground.

In the process of conducting qualitative research, investigators often reach the point where respondents in later rounds of focus groups begin to discuss the same issues that emerged in earlier rounds of focus groups. This is called the saturation point (Glaser & Strauss, 1967). If the goal of a study is to flush out the content domain of a construct, gathering redundant information is not useful. Therefore, once the saturation point is reached, the moderator's guide is changed so that time spent with study subjects can be devoted to uncovering new information.

The focus group data were evaluated using a two-level qualitative data inventory. The first level, which is part of the funnel approach discussed previously, consisted of a more general assessment of the findings. Here, the intent was to sort the data into relatively crude large-level categories that could be used to identify themes to pursue in later rounds of focus groups (Brenner, 1985). Consequently, this first-level assessment of the data was an ongoing process that unfolded over the course of all eight focus groups. The second level of analysis began once the focus groups were complete. Here, members of the research team reviewed the transcripts and independently developed a more detailed coding scheme (Mishler, 1986). Following this, team members met as a group, discussed their coding schemes, and arrived at a consensus on how best to present the data.

The focus groups provided a wealth of information about religion in late life. Two papers were written to explore select areas in detail: One examined negative interaction in the church (Krause, Morgan, Chatters, & Meltzer, 2000a), whereas the other dealt with prayer (Krause, Morgan, Chatters, & Meltzer, 2000b). The findings on prayer will be reviewed briefly to highlight the rich insights that emerged from this phase of the item development strategy.

Most research on prayer focuses on how often people pray and the types of prayers they offer (e.g., prayers of thanksgiving or petitionary prayers requesting specific outcomes; Poloma & Gallup, 1991). However, it became apparent over the course of the focus groups that study participants had a good deal to say about whether prayers are answered, and if they are, how answers are provided. So, for example, some felt that prayers are answered right away, whereas others believe that God answers prayers whenever He is ready. Moreover, some felt they got exactly what they asked for in a prayer. In contrast, others indicated they did not always get what they asked for. However, when these individuals took the time to think about it, they found the answer they got was precisely what they needed most.

This information provides a host of new ways to think about the relationship between prayer and health in late life. For example, older people who expect immediate answers to prayers may become disillusioned, and lose hope, if the anticipated answer is not forthcoming. As research indicates, the loss of hope may be an important risk factor for some health problems (Nunn, 1996). To empirically evalu-

ate this issue, closed-ended questions were crafted in later phases of this study to assess beliefs about the timing of answers to prayers.

A major advantage of using focus groups arises from the fact that it is possible to observe and record how older people talk to each other when they discuss important topics like religion. The words and phrases they use provide excellent raw grist for writing closed-ended question stems. There are, however, several disadvantages in using focus groups. First, respondent burden is heavy because subjects must leave their homes and come to a common location, such as a meeting room in a local hotel. Unfortunately, this makes it difficult for elderly people to attend focus groups if they do not have adequate transportation, or if they are physically challenged. Transportation problems may be dealt with by providing cab rides for those who need them, but little can be done to improve attendance rates among older adults with physical health problems.

The second disadvantage in using focus groups is discussed by Knodel (1995). He argues that the presence of others in the focus group may make it difficult to disclose information of a personal nature. Because some older people may consider their religious beliefs and practices to be a private matter, they may feel uncomfortable expressing their views and feelings in a focus group setting. In view of these limitations, and consistent with the principle of triangulation, the next step in the item development strategy involved conducting a series of one-on-one, in-depth interviews.

In-Depth Interviews

The individual in-depth interviews were conducted face-to-face with study participants in their homes. All interviewers were race-matched with respondents. A new set of subjects was recruited for the in-depth interviews from the HCFA list. Once again, the names of potential respondents were selected using simple random sampling procedures. A total of 131 in-depth interviews were completed successfully. Approximately 61% of these individuals were older White adults, and 39% were elderly African American adults. All interviews were tape-recorded. The interviews typically lasted between 60 and 90 min.

The number of in-depth interviews conducted for this study is unusually large ($N = 131$). In fact, some investigators suggest that as few as eight are sufficient to cover a new domain (McCracken, 1988). However, the members of the research team believed it was important to interview a large number of older people because the content domain of religion is so vast. This decision was ultimately supported by the wealth of data that was obtained during this phase of the study.

Throughout, the funnel approach and the saturation point concepts guided the flow of the in-depth interviews. As a result, the content of the questionnaire changed a number of times as the in-depth interviews were being conducted. It is important to emphasize two points about the way in which the funnel approach was implemented. First, following the procedures used in the focus groups, all in-depth interviews began with general, open-ended questions about religion and concluded with questions about more focused aspects of religion. Second, in-depth interview questions about more

specific aspects of religion were written with input from the focus group data. This means that the use of the funnel technique was extended in this study by carrying insights across two different qualitative methodologies.

During the course of developing the in-depth interview questionnaires, a procedure was implemented that departs significantly from the traditional qualitative approach. As noted earlier when discussing the funnel approach, focused probe questions are typically developed based solely on the input provided by study subjects. However, in the process of writing the in-depth interview questionnaire, information from other sources was taken into consideration. Shortly before the in-depth interviews began, the John Templeton Foundation issued a request for proposals on forgiveness. This initiative was developed in response to a burgeoning literature, which suggests that forgiveness may be an important factor in promoting health and well-being (McCullough, Pargament, & Thoresen, 2000). However, the focus group participants in the present study had very little to say about forgiveness. Even so, the research team decided that since the goal of the project was to develop a comprehensive set of religion measures, it would not be advisable to overlook this potentially important construct. This decision raises a broader question about how to obtain complete coverage of a conceptual domain that has been the subject of empirical and theoretical investigation for some time. For more than 100 years, scholars have been studying religion in an effort to distill its essential elements (James, 1902/1997). This vast literature contains many valuable theoretical insights and important empirical findings (Koenig et al., 2001). We felt it did not make sense to completely disregard this work. Instead, a more profitable approach to designing in-depth interviews involves finding a way to exploit existing material without compromising the inherent advantages associated with this important qualitative methodology.

With this objective in mind, a series of open-ended probe questions on forgiveness were placed at the end of the in-depth interviews. This ensures that the questions on forgiveness did not unduly influence or bias responses to the earlier, more traditional, qualitative probe questions because respondents were given ample time to express their own views before they were presented with questions based on external material. Viewed more generally, this strategy provides a unique way to more tightly integrate qualitative methods and quantitative research findings in the interests of developing the best closed-ended survey items.

As the in-depth interviews progressed, it quickly became evident that the questions on forgiveness evoked the most emotionally charged response that was encountered during the entire study. This was especially true of questions about self-forgiveness. Although data are not available to explain why forgiveness did not arise spontaneously in the focus groups, perhaps the highly personal nature of this topic initially inhibited discussion.

Because so many in-depth interviews were conducted ($N = 131$), the author was the only person to listen to all the tapes. Different members of the research team reviewed segments of the taped interviews dealing with specific topics, such as forgiveness. After these individuals independently coded the themes that emerged from the data, they met with the author

to discuss their findings and arrive at a consensus on the best way to code the data.

The in-depth interviews provided rich insights into how older adults practice religion in their daily lives. The quality of these data may be illustrated by briefly reviewing some of the results on forgiveness. More detailed findings on forgiveness appear in a paper that was written with these data (Krause & Ingersoll-Dayton, 2001). As the in-depth interviews progressed, it quickly became evident that respondents held quite different views about forgiving other people. Although some felt it was their duty as Christians to forgive others automatically, and not require transgressors to do anything first, other study participants believed that transgressors must earn their forgiveness. In particular, this latter group of subjects felt that transgressors must take one or more of the following steps to be forgiven: (1) transgressors must be aware of what they have done; (2) explicitly ask for forgiveness; (3) offer an explanation for the hurtful act; (4) make a resolution not to repeat the offense again; (5) follow-up on this resolution by changing their behavior; and (6) make amends (i.e., provide restitution). These insights set up a series of intriguing questions about the nature of the relationship between forgiveness and health. First, regardless of how it is attained, is forgiveness related to health? Second, does forgiving others without requiring them to do anything first provide the greatest health benefits, or is requiring transgressors to earn forgiveness more beneficial in this respect? Finally, if forgiveness must be earned, do potential health-related benefits increase as transgressors perform progressively more of the steps outlined previously?

In retrospect, it appears as though more valuable data were obtained from the in-depth interviews than from the focus groups. Although it is difficult to support this impression with hard data, study subjects may have provided better information because they felt more comfortable discussing sensitive religious issues in the privacy of the individual in-depth interview (Knodel, 1995). Also, as noted previously, because the in-depth interviews were conducted in the homes of study participants, this methodology made it possible to reach a wider group of older people, including those who were homebound.

Input from Quantitative Studies

The next step in the item development strategy also departs from traditional qualitative procedures. This phase of the study emerged serendipitously. As the focus group and in-depth interviews were being conducted, an opportunity arose to insert closed-ended quantitative questions on religion in a nationwide survey of people affiliated with the Presbyterian Church (USA). Data from the focus groups and in-depth interviews revealed that many older study participants valued church-based social support highly. Although relatively little space was available in the questionnaire for the Presbyterian study, seven closed-ended items were crafted from the qualitative data to assess three aspects of church-based support: Emotional support from rank-and-file church members, emotional support from the pastor, and spiritual support from church members. Spiritual support involves things like sharing religious experiences and helping others live according to their religious beliefs. Data from the nationwide

Presbyterian survey provided an excellent opportunity to gain a preliminary sense of the quality of the newly developed items. In particular, a series of analyses were performed ranging from an examination of simple frequency distributions to the estimation of a sophisticated second-order confirmatory factor model. The results of these analyses appear in a paper by Krause, Ellison, Shaw, Marcum, and Boardman (2001).

The Presbyterian survey was useful for several reasons. First, the confirmatory factor model revealed that the newly developed measures had good psychometric properties (i.e., good item and scale reliability).

Second, the Presbyterian study provided an opportunity to conduct a preliminary assessment of the construct validity of the newly devised, church-based support items. Construct validity is assessed by seeing whether a new set of items is related to an established outcome measure in a theoretically meaningful way. The Presbyterian survey also contained a set of widely used indicators on religious coping responses that were developed by Pargament (1997). These items assess specific ways that people use religion when they are confronted by stressful events. Based on the premise that the selection of religious coping responses are socially determined, Krause and colleagues (2001) examined whether the three dimensions of the church-based support discussed previously are related to religious coping. Data indicate that people are especially likely to use religious coping responses when they receive spiritual support from church members. Even though emotional support from the pastor was also associated with greater use of religious coping methods, the relationship was not as strong. In contrast, emotional support from church members had no effect on religious coping.

Third, findings from the empirical work with the Presbyterian data also provided valuable information on how to implement the funnel approach in the in-depth interviews. As discussed earlier, the funnel technique involves using data from earlier rounds of in-depth interviews to develop more focused questions for later rounds of in-depth interviews. The decision to devote scarce interview time to probe specific areas in detail is based on the nature of the themes that emerge from the initial in-depth interviews. However, if the goal of a study is to ultimately develop closed-ended quantitative survey measures, then relatively unique criteria must be used for identifying themes that should be probed more deeply. In particular, an investigator must determine whether different themes that emerge from the qualitative data will lead to the development of different quantitative scales, or whether the themes assess content areas that are so closely related that they can be more economically assessed with a single scale. It is sometimes difficult to make this decision based on the qualitative data alone. This was true with respect to the church-based social support measures identified previously. In particular, it was hard to imagine how rank-and-file church members could provide spiritual support to respondents without also giving them emotional support at the same time. However, the empirical analysis of the Presbyterian data suggests that even though the two dimensions of church-based support are correlated, spiritual support is related to religious coping, whereas emotional

support from church members is not. This differential impact suggests that the two dimensions of church-based support are conceptually distinct and that it would be useful to continue to invest in-depth interview time in exploring them both. Viewed from this perspective, the analysis of the Presbyterian data provides a relatively unique way of showing how both qualitative and quantitative methods can be used simultaneously to develop closed-ended survey items.

Developing Preliminary Quantitative Measures

Once the in-depth interviews were complete, a set of closed-ended survey questions was developed. These questions came from three sources. First, some questions were taken from existing scales (e.g., measures of religious commitment, as well as some religious coping items). Second, measures devised by other investigators were modified on the basis of information gleaned from the focus groups and in-depth interviews (e.g., some indicators of emotional support from church members). Finally, because good measures could rarely be found in the literature, the bulk of the items were developed from scratch.

It is important to provide more detail on how new survey items were written when good indicators could not be found elsewhere. Two issues figured prominently in this process. The first involved creating a list of all facets of religion that emerged from the focus groups and in-depth interviews. The two-level qualitative inventory discussed earlier was invaluable in this respect.

The second key issue involved confronting the problem of depth versus breadth. Data obtained in the focus groups and in-depth interviews were incredibly rich, and the number of questions that could be written from this material seemed almost limitless. However, because the final step in the item development strategy consisted of a nationwide survey lasting 70 minutes, crucial decisions had to be made about how many dimensions of religion to cover, and how many items to devise for each dimension. There does not appear to be much guidance in the literature on how to make these decisions. The following approach was used in this study.

First, the members of the research team estimated the total number of questions that could be administered in a 70-min interview. Second, in estimating the likely number of items, team members had to take into account the fact that questions would also be asked about health, psychological well-being, and a number of factual matters, such as age, education, and marital status. Once the amount of questionnaire space that could be used to measure all areas of religion was established, the number of specific dimensions that could be covered was estimated by turning to the work of Andrews (1984). His sophisticated latent variable modeling analyses reveal that the ideal length of a scale should be between two and four items. The upper limit of four items was used in the present study. Although departures from the target of four indicators ultimately arose, this general guideline was, nevertheless, useful. Finally, once an estimate was derived of the number of dimensions of religion that could be covered, difficult decisions had to be made about what to include. This was clearly the most challenging part of this study phase. Because the ultimate goal of this study was to examine the relationship between religion and health, decisions

were based, in part, on whether a particular dimension of religion is likely to affect the health of older people. Dimensions of religion that might be related to health were identified by turning to the literature, as well as insights gleaned from the focus groups and in-depth interviews.

Panel of Experts

Seven scholars with outstanding reputations in the area of religion graciously agreed to review the items that were developed for this study. The following researchers were members of this group: Linda Chatters, Christopher G. Ellison, Ellen Idler, Harold G. Koenig, Jeffrey S. Levin, Kenneth I. Pargament, and Robert Taylor. It should be emphasized at the outset that even though these individuals provided invaluable input, the author is solely responsible for any problems with the closed-ended survey items that were developed in this study. Although expert panels have been used to develop items in other studies, a protocol for fully exploiting the expertise of these individuals is difficult to find in the literature. The following procedures were developed especially for this study.

First, a complete draft of all the religion items was sent to each of the experts. They were instructed to rate each indicator on a scale that ranged from 1 to 5, where 5 meant the question stem and the item response categories were of high quality, and a score of 1 denoted items that were most in need of revision. These data were mailed back to the study field office where frequency distributions of the ratings for each question were tabulated.

Following this, the experts were all flown to Ann Arbor for an intensive 2-day review of the study measures. The ratings provided before the meeting were used to structure the discussions. In particular, the meeting began by focusing on items with the least favorable ratings. The problems with each measure were identified, discussed, and potential solutions were proposed. The 2-day session with the experts was tape-recorded and transcribed.

An example may help illustrate how the expertise of the panel members was put to good use. Recent research suggests that, as people grow older, they turn control over specific domains in their lives to trusted others (Schulz & Heckhausen, 1996). For example, some older people turn over the management of their financial affairs to a grown offspring. During the course of the focus groups and in-depth interviews, the subjects in our study indicated they turn control over to God as well. A series of preliminary items were devised to capture this domain. Initially, these items focused on turning things over completely to God, but the group of experts recommended that a second approach should also be considered. In particular, they argued that people may only partially rely on God while continuing to exercise control themselves. Items were, therefore, developed to capture this form of collaborative control with God. Following this two-part strategy makes it possible to see if complete control by God, or collaborative control with God, is most likely to enhance the health and well-being of older people.

In addition to reviewing question stems and response formats, the expertise of the group was also used to help determine the order of the closed-ended questions in the interview schedule. Toward the end of the 2-day meeting, the experts

were each given a stack of index cards. The name of one dimension of religion was written on each card. The experts were asked to sort the cards according to where the questions should appear in the final closed-ended questionnaire. A consensus was determined by tabulating the rankings provided by the group.

Based on the input from the group of experts, a number of changes were made to the newly devised religion questions. The tapes and transcripts of the meeting with the experts proved to be quite valuable in this respect. In the process, problems and revisions suggested by these experts were checked against the focus group and in-depth interview data.

Cognitive Interviews

Once a good set of preliminary closed-ended questions was in place, a series of cognitive interviews were conducted. Cognitive interviews involve presenting study subjects with a closed-ended question followed by a series of open-ended probes. The purpose of the open-ended probe questions is to see if study subjects understand the closed-ended question in the intended manner, to see if respondents can provide a better way of phrasing question stems, and to see if subjects feel comfortable with the closed-ended response options.

Eighty-five cognitive interviews were conducted with a new sample of subjects. Once again, these study participants were selected from the HCFA list, using simple random sampling procedures. Up to this point, older adults were excluded from the study if they indicated that religion was not at all important to them. However, this exclusion criteria was relaxed at this point, and all elderly people were included in the remaining steps of this study regardless of how important religion may be to them. Forty-five of the participants in the cognitive interviews were older White adults, and 40 were elderly African American adults. All cognitive interviews were conducted face-to-face in the homes of the study participants.

The closed-ended religion items were broken down into four blocks. Cognitive interviews were conducted using approximately 20 subjects for each of the four blocks of questions. It was not necessary to probe some closed-ended questions because a good deal is already known about them (e.g., "At the present time, what is your religious preference?"). The cognitive interviews lasted between 60 and 90 min. All cognitive interviews were tape-recorded.

Initially, members of the research team were concerned that study participants would quickly tire when they found that each closed-ended question would be followed by a series of open-ended probes. This did not prove to be the case. Instead, the older people in this study worked hard to help the research team improve the closed-ended questions. This high level of motivation was due in part to the way the cognitive interviews were introduced. Subjects were told that a great deal of time had been spent speaking with older people about religion in an unstructured manner. They were also told that a series of questions had been written about religion based on these conversations, but to know if these items are any good, their critical input was essential. Subjects were asked to work hard to help the research team uncover problems and find new ways to improve the questions. This approach represents a modified version of Cannell's strategy for

increasing respondent motivation and commitment (Cannell, Miller, & Oksenberg, 1981).

Two broad approaches may be followed in developing probe questions for cognitive interviews. The first is to ask very general follow-up questions, as in the think-aloud strategy (see Foddy, 1998, for a recent discussion of this technique). An example will help clarify the nature of these general probe questions. Using religion to derive a sense of meaning in life emerged as an important theme in this study. Consequently, a series of closed-ended questions were devised to assess this domain. One item from this battery asked subjects if they felt that, "God has a specific plan for my life." This indicator was followed with a broad probe: "Can you tell me a little more about what you were thinking about when you answered this question?"

These general probe questions did not appear to be especially useful. Instead, a second approach involving more focused probe questions was more helpful. This strategy can be illustrated by turning to forgiveness. Some subjects in the qualitative studies indicated that when they were hurt by someone, they were usually able to "forgive and forget." A closed-ended question was written to assess this topic. After administering the closed-ended question, subjects were asked the following focused probe questions: "What does the phrase 'forgive and forget' mean to you? If you were to ask a friend about this, how would you do it—what words would you use to see if they can forgive and forget?" The observation that focused probe questions tend to produce better feedback than general probe questions is consistent with recent research on cognitive interviewing by Foddy (1998).

An example may help clarify how the focused cognitive interview probes were used to revise the religion measures. In his outstanding work on the measurement of religious coping, Pargament (1997) devised an item that asks respondents how often they look to God for strength and guidance in a crisis. The respondents in our focus groups and in-depth interviews indicated they turned to God for this purpose, but the members of the research team were concerned that the item devised by Pargament (1997) was double-barreled because "strength" and "guidance" may not mean the same thing to study participants. To evaluate this possibility, participants in the cognitive interviews were presented with the original item developed by Pargament (1997). Focused probe questions were subsequently asked to see if they felt that turning to God for strength was the same as turning to God for guidance. The wide majority of study subjects indicated they were different and recommended that separate questions be asked about these issues. Based on their input, two separate items were developed: "I look to God for strength in a crisis," and "I look to God for guidance when difficult times arise."

The cognitive interviews were also used to evaluate the closed-ended responses for the religion items. For example, subjects were asked whether they felt it is easier to answer a question using a standard five-point Likert scale, or whether selecting a response from a Cantril ladder (scored 1–10) made more sense to them. An effort was also made to learn more about how older people calibrate closed-ended responses. For example, subjects were asked whether they used a particular religious coping response a great deal,

some, only a little, or not at all. Cognitive interview probe questions were subsequently asked to determine how much difference (if any) there was between a response of "some" and a response of "only a little."

The newly developed religion items were revised again after members of the research team listened to the tapes of the cognitive interviews. These revisions were made by consulting the data gathered from the focus groups and in-depth interviews, and by turning to the feedback provided by respondents during the cognitive interviews.

Pilot Study

The closed-ended religion items were administered to a new sample of 98 older subjects. Half the respondents were older White adults ($n = 49$), and half were elderly African Americans ($n = 49$). These subjects were, once again, selected from the HCFA list. All pilot test interviews were conducted face-to-face in the homes of the study participants. Only closed-ended quantitative questions were administered in this phase of the study. As a result, this step in the item development strategy more closely approximates a traditional pilot test.

The goal of the pilot study was to check the length of the survey, to examine frequency distributions to make sure the indicators had sufficient variance, and to perform exploratory factor analyses to examine the structure and psychometric properties of the newly developed scales. This information was supplemented with feedback from the interviewers about any problems they encountered with the questionnaire. Some minor revisions were made in some of the questions at this point. In addition, the exploratory factor analyses helped identify a few items that could be eliminated from some scales, thereby holding down the overall length of the interview schedule.

After this phase of the item development strategy was complete, a total of 175 closed-ended questions were in hand that assessed 14 major dimensions of religion. Copies of the religion items are available on request from the author.

Nationwide Survey

The next step in the item development strategy involved administering the closed-ended questionnaire to a nationwide sample of 1,500 older people. Data were collected by Harris Interactive (formerly Louis Harris and Associates). Interviewing was completed in August 2001. The sample consisted of 750 elderly White subjects and 750 elderly African Americans. The sampling frame for the study was again provided by HCFA. Greater detail on the sampling procedures is available from the author.

Before the nationwide survey went into the field, the questionnaire was reviewed closely by staff members at Harris Interactive who have considerable experience in writing closed-ended survey questions. A few very minor revisions were made to the religion items. Then the entire interview schedule was evaluated in two brief pilot tests, consisting of 15 subjects each.

Psychometric Testing

The final step in the item development strategy is currently in progress. A series of substantive papers are being

written to explore the relationships between select dimensions of religion, health, and well-being (copies of these papers are available from the author). Detailed psychometric testing is being performed on the religion measures that are used in each paper. Four issues involving these tests are discussed briefly below.

First, all multiple item scales are examined with both exploratory and confirmatory factor analyses. The goal is to see whether items that were designed to capture a specific dimension of religion cluster together, and whether the indicators measure each factor well (i.e., whether the factor loadings are equal to or greater than .400).

Second, the findings from the confirmatory factor analyses are being used to compute internal consistency reliability estimates for each scale. More specifically, based on a formula provided by Rock, Werts, Linn, and Jöreskog (1977), the factor loadings and measurement error terms associated with each item are used to derive reliability estimates. So far, the reliability estimates for the newly devised scales are generally in excess of .800.

Third, in the process of performing the confirmatory factor analyses, tests are being performed for measurement invariance (Bollen, 1989). As noted earlier, older White subjects make up half the sample obtained in the nationwide survey, whereas the other half consists of older Black subjects. Consequently, it is important to see whether the factor loadings and measurement error terms are equivalent in both racial groups. If these parameter estimates are the same, it is reasonable to conclude that the survey items mean the same thing to older White respondents and older Black respondents. One goal in conducting the focus groups, in-depth interviews, and cognitive interviews for this study was to develop survey items that mean the same thing to older White and older Black people. However, the assessment of measurement equivalence in this context is largely based on subjective evaluations of feedback provided by the focus group, in-depth interview, and cognitive interview participants. The empirical tests of measurement equivalence that are now being performed with the nationwide data are a nice complement to these qualitative strategies, and provide yet another way of more tightly merging qualitative and quantitative methods in studies designed to develop closed-ended survey measures.

Finally, in addition to assessing issues in measurement invariance and scale reliability, it is also important to evaluate the validity of the newly devised measures. Unfortunately, it is difficult to estimate some types of validity (i.e., predictive validity) with data that have been gathered at one point in time only (see Carmines & Zeller, 1979, for an insightful discussion of scale validity). However, it is possible to evaluate construct validity with the data on hand. As noted earlier, construct validity is evaluated by embedding new measures in substantive conceptual models to see whether the new indicators are related to select outcomes in theoretically meaningful ways. Although the establishment of construct validity is an ongoing task that requires replication across a series of studies, the results that have emerged from the analyses that have been done so far are very encouraging. For example, as one might anticipate, older people who receive spiritual support from their fellow parishioners indicate they

have a closer relationship with God, and people who feel closer to God report they are more hopeful about the future (further detail on this is available from the author).

RESULTS AND DISCUSSION

Some time ago, Bohrnstedt (1983, p. 69) wrote that, "Measurement is a sine qua non of any science." This makes sense because the relationship between two or more measures cannot be evaluated properly until good indicators of each construct are firmly in place. Unfortunately, it seems that researchers spend far more time discussing how data are analyzed, whereas far less attention is devoted to how study measures were devised. This problem arises, in part, because sound protocols for developing closed-ended survey items have rarely appeared in the literature. A basic premise in the present study is that merging qualitative and quantitative methods represents a promising way to approach this problem. Some time ago, Lazarsfeld (1944) recommended that qualitative and quantitative methods be used for this purpose, but there is relatively little concrete guidance on how to implement this strategy in practice. The purpose of this study was to provide a practical, yet comprehensive, strategy for developing closed-ended survey items. Beginning with focus groups, and culminating in the psychometric testing of data obtained in a nationwide probability sample of older people, the intent was to highlight problems and provide helpful solutions for those wishing to study the content domain of a wide range of constructs. Although these procedures have all been examined previously in the literature, there do not appear to be any studies that pull them together in one place.

A tremendous amount of work was needed to execute all nine steps in this study. In fact, it took three full years to complete this task. Not counting the nationwide survey, the research team spoke with 399 older adults. In addition to providing vitally important information on how to develop closed-ended survey questions, the qualitative interviews triggered a flood of new ideas about theoretical or substantive issues in the field of religion and aging. As discussed earlier, a great deal was learned about the steps that may be followed in the process of forgiveness, as well as beliefs about how prayers are answered.

Those wishing to use the item development strategy outlined above should pay attention to the shortcomings in this approach. Three limitations are discussed briefly below. First, because merging qualitative and quantitative methods is costly and labor-intensive, it is incumbent upon those who advocate this approach to demonstrate that this strategy produces results that are superior to findings obtained with more traditional approaches to survey item development (e.g., relying on a panel of experts alone). Moreover, we need to devise ways of seeing whether each step in the item development strategy outlined earlier has an equal payoff, or whether one step (e.g., focus groups) is less effective than another (e.g., in-depth interviews). Unfortunately, protocols do not appear to exist for tackling these difficult tasks. Developing a feasible evaluation component should be a top priority for investigators wishing to merge qualitative and quantitative research methods in the same study.

Second, with the exception of the nationwide survey, data

for all steps in the item development strategy were gathered in Washtenaw County, MI. This was done because practical, as well as economic, considerations made it difficult to conduct focus groups and in-depth interviews in a wider geographical area. Nevertheless, research indicates that there are regional differences in the way Christianity is practiced (Hunt & Hunt, 2001). Consequently, the items developed in the present study may not fully capture how older people in other regions of the nation, such as the rural southeast, view their faith.

Third, as noted earlier, many interesting dimensions of religion emerged from the qualitative interviews, but questions were not written for all of them. So, for example, a good deal of discussion in the in-depth interviews involved changes that older adults experienced in the way they practiced religion over the course of their lives. In fact, a separate paper was written on this issue (Ingersoll-Dayton, Krause, & Morgan, in press). Even so, questions on change in religion over the life course were not included in the final nationwide survey. As noted earlier, there are no firm guidelines for figuring out which domains should be pursued and which should be eliminated. In the end, this decision was based on a subject judgment concerning which domains are most likely to affect health in late life. However, researchers who are interested in relating religion to other outcomes, and investigators who treat religion as a dependent variable, may find that the dimensions of religion that were excluded from the present study are vitally important for their purposes.

Both qualitative and quantitative researchers may question some of the procedures that were used in the item development strategy. For example, qualitative investigators may not feel comfortable with feeding empirical findings from ongoing quantitative studies into the in-depth interviews (see step 3 in Table 1). Similarly, as discussed previously, quantitative researchers may feel uneasy about recruiting subjects for focus groups and in-depth interviews from a local geographical area. Consequently, the greatest contribution of the present study may arise from the fact that it provides a concrete forum for opening a more focused dialogue on how to best develop closed-ended survey items for studies of older adults.

ACKNOWLEDGMENTS

This research was supported by the following grant from the National Institute on Aging: RO1 AG14749 (Neal Krause, Principal Investigator).

Address correspondence to Dr. Neal Krause, Department of Health Behavior and Health Education, School of Public Health, the University of Michigan, Ann Arbor, MI 48109-2029. E-mail: nkrause@umich.edu

REFERENCES

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 48, 391–424.
- Bohrnstedt, G. W. (1983). Measurement. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 69–121). New York: Academic Press.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brenner, M. (1985). *The research interview: Uses and approaches*. London: Academic Press.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389–437). San Francisco: Jossey-Bass.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Chatters, L. M. (2000). Religion and public health: Public health research and practice. In J. E. Fielding, L. B. Lave, & B. Stanfields (Eds.), *Annual review of public health* (Vol. 21, pp. 335–367). Palo Alto, CA: Annual Reviews.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257–282). New York: Russell Sage Foundation.
- Denzin, N. K. (1970). *The research act: A theoretical introduction to sociological methods*. Chicago: Aldine.
- Ellison, C. G., & Levin, J. S. (1998). The religion-health connection: Evidence, theory, and future directions. *Health Education & Behavior*, 25, 700–720.
- Faust, D., & Minor, R. A. (1986). The empiricist and his new clothes: DSM-III in perspective. *American Journal of Psychiatry*, 143, 962–967.
- Fetzer Institute and the National Institute on Aging. (1999). *Multidimensional measurement of religiousness/spirituality for use in health research*. Kalamazoo, MI: Fetzer Institute.
- Foddy, W. (1998). An empirical evaluation of in-depth probes used to pretest survey questions. *Sociological Methods and Research*, 27, 103–133.
- Fultz, N. F., & Herzog, A. R. (1993). Measuring urinary incontinence in surveys. *The Gerontologist*, 33, 708–713.
- Gallup, G. H. (1999, July 14). *The age factor in religious attitudes and behavior* [On-line]. Retrieved December 2001 from www.gallup.com/poll/releases
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Hunt, L. L., & Hunt, M. O. (2001). Race, region, and religious involvement: A comparative study of Whites and African Americans. *Social Forces*, 80, 605–631.
- Ingersoll-Dayton, B., Krause, N., & Morgan, D. L. (in press). Religious trajectories and transitions over the life course. *International Journal of Aging and Human Development*.
- James, W. (1902/1997). *Selected writings—William James*. New York: Book-of-the-Month Club.
- Knodel, J. (1995). Focus groups as a qualitative method for cross-cultural research in social gerontology. *Journal of Cross-Cultural Gerontology*, 10, 7–20.
- Koenig, H. G. (1994). *Aging and God: Spiritual paths to mental health in midlife and later years*. Binghamton, NY: Haworth.
- Koenig, H. G., McCullough, M. E., & Larson, D. B. (2001). *Handbook of religion and health*. New York: Oxford University Press.
- Kozma, A., & Stones, M. J. (1987). Social desirability in measures of subjective well-being: A systematic evaluation. *Journal of Gerontology*, 42, 56–59.
- Krause, N., Ellison, C. G., Shaw, B. A., Marcum, J. P., & Boardman, J. (2001). Church-based support and religious coping. *Journal for the Scientific Study of Religion*, 40, 637–655.
- Krause, N., Ellison, C. G., & Wulff, K. M. (1998). Church-based emotional support, negative interaction, and psychological well-being: Findings from a national sample of Presbyterians. *Journal for the Scientific Study of Religion*, 37, 725–741.
- Krause, N., & Ingersoll-Dayton, B. (2001). Religion and the process of forgiveness in late life. *Review of Religious Research*, 42, 252–276.
- Krause, N., Morgan, D. L., Chatters, L. M., & Meltzer, T. (2000a). Negative interaction in the church: Insights from focus groups. *Review of Religious Research*, 41, 522–545.
- Krause, N., Morgan, D. L., Chatters, L. M., & Meltzer, T. (2000b). Using focus groups to explore the nature of prayer in late life. *Journal of Aging Studies*, 14, 191–212.
- Lazarsfeld, P. F. (1944). The controversy over detailed interviews: An offer for negotiation. *Public Opinion Quarterly*, 8, 38–60.
- Luborsky, M. R., & Rubinstein, R. L. (1995). Sampling in qualitative research. *Research on Aging*, 17, 89–113.
- Madill, A., Jordan, A., & Shirley, C. (2000). Objectivity and reliability in

- qualitative analysis: Realist, contextualist, and radical reconstructionist epistemologies. *British Journal of Psychology*, 91, 1–20.
- May, K. (1991). Interview techniques in qualitative research: Concerns and challenges. In J. M. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 188–201). Newbury Park, CA: Sage.
- McCracken, G. (1988). *The long interview*. Thousand Oaks, CA: Sage.
- McCullough, M. E., Pargament, K. I., & Thoresen, C. E. (2000). *Forgiveness: Theory, research, and practice*. New York: Guilford.
- Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.
- Morgan, D. L. (1988). *Planning focus groups*. Thousand Oaks, CA: Sage.
- Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research*, 8, 362–376.
- Morgan, D. L., & Krueger, R. A. (1998). *The focus group kit*. Thousand Oaks, CA: Sage.
- Nunn, K. P. (1996). Personal hopefulness: A conceptual review of the relevance of the perceived future to psychiatry. *British Journal of Medical Psychology*, 69, 227–245.
- Oksenberg, L., Cannell, C., & Kalton, G. (1991). New strategies for pre-testing survey questions. *Journal of Official Statistics*, 7, 349–365.
- Pargament, K. I. (1997). *The psychology of religion and coping: Theory, research, and practice*. New York: Guilford.
- Poloma, M. M., & Gallup, G. H. (1991). *Varieties of prayer: A survey report*. Philadelphia: Trinity Press.
- Princeton Religion Research Center. (1994). *Religion in America (1994 Supplement)*. Princeton, NJ: Gallup Poll.
- Rock, D. A., Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1977). A maximum likelihood solution to the errors in variables and errors in equations model. *Journal of Multivariate Behavioral Research*, 12, 187–197.
- Schulz, R., & Heckhausen, J. (1996). A life span model of successful aging. *American Psychologist*, 51, 702–714.

Received December 10, 2001

Accepted February 27, 2002

Decision Editor: Fredric D. Wolinsky, PhD