Focus Groups in Psychological Assessment: Enhancing Content Validity by Consulting Members of the Target Population.

3 authors, including:

Dawne Vogt
U.S. Department of Veterans Affairs
84 PUBLICATIONS   **3,507** CITATIONS

SEE PROFILE

Daniel W King
Boston University
154 PUBLICATIONS   **9,827** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Longitudinal study of veterans' well-being as they transition from military service View project

Project   SHARP - Stress Health, and Aging Research Program View project

Focus Groups in Psychological Assessment: Enhancing Content Validity by Consulting Members of the Target Population

Dawne S. Vogt, Daniel W. King, and Lynda A. King
Veterans Affairs Boston Healthcare System and Boston University

A review of articles in *Psychological Assessment* reveals that many researchers develop instruments without the benefit of consultation with members of the target population. To the extent that researchers do consult the target population, most fail to bring consultation in early enough to inform the identification and specification of key constructs. Moreover, this consultation typically takes the form of one-to-one interviews. The authors' goal in this article was to elaborate on the importance of population consultation as part of content validation and to critically evaluate the potential of one method of consultation, focus groups, to inform multiple stages of instrument development. The authors suggest that this method holds promise for enhancing the content validity of instruments and, ultimately, the validity of research findings.

All researchers are concerned with the extent to which the tools or assessment devices they use measure the constructs they are intended to measure. The degree to which they do has significant implications for the validity of conclusions drawn from research results, as well as the appropriateness of applications to practice and treatment (Haynes, Richard, & Kubany, 1995). On the one hand, the use of assessment instruments that do not adequately represent constructs may threaten the validity of research findings. On the other hand, the use of instruments that are content valid for their intended purpose, and that possess other requisite psychometric properties, can lead to advances in theory and practice.

Haynes et al. (1995) advocated for the application of a multimethod approach to enhancing the content validity of newly developed instruments. They described several methods that researchers have used as part of their content validation procedures. Among these, they identified consultation with members of the target population (i.e., those for whom the instrument is intended) as one of the least frequently applied methods. Our goal in the present article is to build on and extend the work of Haynes et al. by elaborating on the importance of consultation with members of the target population as part of content validation, and critically evaluating the potential of one method of consultation—focus groups—to inform the conceptualization of key constructs and item development.

First, a discussion of content validity and its relevance to the instrument development process is provided. Next, we briefly review commonly used content validation procedures and present the focus group methodology as an attractive alternative to one-to-one interviews for the purpose of population consultation. Strengths and limitations of the focus group methodology are then described. After this, we provide an example of how focus groups were applied in one instrument-development project and present evidence for the usefulness of this approach. We close with several general principles and recommendations for researchers who wish to apply a similar procedure in their own work.

Content Validity as Part of Construct Validation

According to classical test theory, support for a hypothesized construct is garnered through the process of construct validation. This process includes (a) theory development, evaluation, and expansion and (b) the identification, construction, and refinement of indicators of the construct as it is embedded in the theoretical system (L. A. King & King, 1990). There are four main activities that have been identified in this process: (a) *definition*, description of a construct and how it operates within a theoretical network; (b) *operationalization*, translation of the construct into observable behaviors or events; (c) *evaluation of convergent and discriminant validity*, examination of relationships with indicators of conceptually similar and dissimilar constructs; and (d) *evaluation of nomological validity*, the assessment of relationships within the theoretical system to which the construct belongs (Campbell, 1960; Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Loevinger, 1957; Nunnally & Bernstein, 1994; Schwab, 1980). As L. A. King and King (1990) noted, these components interrelate and build on one another.

A critical aspect of the construct validation process involves assessing and ensuring the content validity of instruments used to

measure key constructs (Messick, 1975). Content validity is especially relevant to the first two components of the construct validation process: the definition and operationalization of the construct. *Content validity* is defined by "the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose" (Haynes et al., 1995, p. 238). The elements of a questionnaire include individual items, response formats, and instructions. This article focuses on the content validity of items: a full explication of content validity as it relates to other elements of the assessment instrument is available in Haynes et al. (1995). The *relevance of an assessment instrument* refers to the appropriateness of its elements for assessing a given construct. For example, a critical issue of relevance regards the extent to which items tap the construct of interest. The *representativeness of the instrument* refers to the extent to which its elements are proportional to the facets of the key construct. This has also been referred to as "the epistemic link" or "the rule of correspondence" (Nunnally & Bernstein, 1994). A measure that includes items that assess, proportionally, all critical domains of the construct and does not include other items outside the construct can be considered to be representative of that construct. Of course, an assessment instrument can appear to be representative but contain irrelevant items. Likewise, an instrument may contain relevant items but not tap proportionately all facets of a target construct, and thus, may not meet the criterion for representativeness.

Although some researchers have treated content validity as just another aspect of the construct validation process, others have argued that it is a prerequisite for establishing other types of validity. According to this latter perspective, evidence for additional aspects of construct validity (e.g., convergent or discriminant validity) for a measure that is not relevant to and representative of the key construct is difficult, if not impossible, to interpret (e.g., Switzer, Wisniewski, Belle, Dew, & Schultz, 1999). At the very least, the content validity of instruments sets the upper limit for construct validity (Haynes, Nelson, & Blaine, 1999). In other words, from a rational test development perspective, findings regarding other forms of validity (e.g., convergent and discriminant validity, predictive validity, concurrent validity) are applicable to a target construct only to the extent that the instrument is actually assessing the construct.

As noted previously, the use of assessment instruments that lack content validity can lead to invalid conclusions as well as the misapplication of findings to practice and treatment (Haynes et al., 1995). To illustrate this point, imagine a study of the impact of workplace sexual harassment on job satisfaction that involves administering a measure of global job satisfaction. Imagine further that this instrument omits the domain of interpersonal relationships (e.g., "satisfaction with one's supervisor," "satisfaction with co-worker relations"), focusing instead on other aspects, such as "perceptions of work as interesting and challenging" and "satisfaction with pay and benefits." On the basis of this omission, researchers might draw the erroneous conclusion that experiences of workplace sexual harassment are unrelated to job satisfaction, when in fact, there may be a meaningful association with interpersonal satisfaction that goes undetected because of the poor content validity of the instrument intended to assess global job satisfaction.

It is important to note that inferences about the content validity of an instrument are conditional. Assessment instruments may have different functions and may be content valid for some functions but not for others (Haynes et al., 1995; Mitchell, 1986). For example, a measure may be content valid for the purpose of screening for anxiety but invalid for the purpose of diagnosing anxiety. The same measure may be valid for assessing situational or state anxiety but invalid for assessing chronic or trait anxiety. Likewise, a measure may be content valid for use with one population but not with another (Haynes et al., 1995; Nunnally & Bernstein, 1994; Suen, 1990). For example, a measure of job-related stressors for higher level salaried personnel might not be suitable for use with nonsupervisory hourly workers. Also, a measure of war-related stressors that is valid for Vietnam veterans may not be valid for veterans of more contemporary military deployments, such as those deployed to the Middle East and Southwest Asia (e.g., Iraq, Afghanistan). Moreover, content validity is a dynamic aspect of an instrument. As constructs are refined over time, the content validity of instruments intended to assess them will necessarily decline.

Content Validation Procedures

Content validation has been conceptualized as a multimethod, quantitative, and qualitative process that is applicable to all elements of an assessment instrument (Haynes et al., 1995). Content validation includes, but is not limited to, careful specification of constructs, review of scaling procedures by content-validity judges, and consultation with experts and members of the population. As Haynes et al. (1995) noted in their review of content validation procedures in the literature, there is a need for improvement in the content validation procedures typically used by researchers. This need is further underscored by evidence for the poor content validity of instruments used in numerous areas of study. For instance, and as will be discussed in greater detail shortly, researchers who study war-related stressors have been criticized for relying on outdated, inappropriate (and thus, content-invalid) measures (Marshall, Davis, & Sherbourne, 1999). Likewise, researchers who study rape-related posttraumatic stress disorder have been encouraged to develop instruments that are more content valid for women who are victims of sexual assault (e.g., more sensitive, behaviorally defined screening questions for the assessment of rape experiences; Resnick, Kilpatrick, & Lipovsky, 1991). Researchers involved in the assessment of psychopathology also have been criticized for using the same measures to assess different cultural groups and, in particular, for adopting those measures developed for use with the dominant culture for the purpose of assessing other cultures (Marsella & Kameoka, 1989). It is likely that content-validity problems arise in other areas of study as well.

Consultation With Members of the Target Population

As mentioned previously, one method of content validation involves consulting with experts or members of the target population. In fact, expert consultation is considered by some to be the sin qua non of content validity (Messick, 1995). Although the term *expert* has typically been used to refer to researchers who are knowledgeable in the specific topic area, members of the population under study may also be considered "experts" in some cases. For instance, in the study of war-related stressors, veterans who have had these experiences may be the best source of information

regarding how they should be conceptualized. Members of the population may also provide useful input at the item-development stage, as they can review items for their ease of understanding and relevance to and representativeness of the construct.

Although some authors have addressed the importance of consulting members of the population in the identification and specification of constructs (e.g., DeVellis, 1991; Fagot, 1992; Frank-Stromborg, 1989; Haynes et al., 1995; Kubany et al., 1996; Loo et al., 2001; Lynn, 1985; Van Gerwen, Spinhoven, Van Dyck, & Diekstra, 1999), researchers rarely follow this advice (Haynes et al., 1995; Smith & McCarthy, 1995). According to Haynes et al. (1995), carefully structured, open-ended interviews with members of the target population can increase the chance that items are content valid for their intended purpose and can also suggest additional facets and the need for construct refinement. We elaborate on this point by further suggesting that consultation with members of the target population can—and should—be used to inform the initial identification and specification of key constructs. Moreover, we argue that the focus group methodology, an approach that involves moderator-facilitated discussions among multiple participants, may be especially helpful in this regard.

Review of Psychological Assessment Articles

A review of articles published in *Psychological Assessment* in the 7-year period following Haynes et al.'s (1995) publication suggests that consultation with members of the population continues to be rarely used by authors presenting new instruments in this journal. Almost half of the authors (12 of 26) who reported developing items for a new questionnaire did not incorporate consultation with members of the population in their content validation procedures. Only about 1 in 4 (27%) consulted members of the population for the purpose of construct identification and specification; the majority of researchers who consulted members of the population (73%) did so for the purpose of item review and refinement. Although such consultation is certainly appropriate at that stage of scale construction, the failure to bring consultation in early enough to inform the identification and specification of key constructs may result in conceptualizations that are faulty and items that do not address important facets of constructs. Omissions of this nature are likely to go unnoticed at later stages of scale construction.

Moreover, to the extent that researchers did report consulting members of the population in the process of instrument development, this consultation typically took the form of one-to-one interviews and rarely drew from the potential of focus group methodology. Of the articles published in this 7-year period, only one (Malgady, Rogler, & Cortes, 1996) indicated using focus groups in the early stage of instrument development. In this study, focus groups with members of the Puerto Rican community were used to inform both the conceptualization and operationalization of idioms of distress. In turn, a scale composed of identified idioms demonstrated incremental validity for predicting clinical status. That is, the inclusion of the measure that was informed by focus groups predicted clinical status above and beyond standard measures of symptomatology.

As is evident from this review, consultation with members of the target population for the purpose of conceptualizing key constructs, and in particular, consultation in the form of focus groups, is vastly underused by researchers presenting new instruments in

Psychological Assessment. To the extent that this is true among articles published in this well-respected journal, it is even more likely the case for instruments introduced in less prestigious journals. In fact, our broader review of the literature revealed a limited number of instrument development projects that were informed by focus groups. On the basis of a search of journal articles that appeared in PsycINFO between 1967 and 2002, and that used the keywords *focus groups* and either *instrument* or *measure*, we identified only 23 scale construction endeavors that drew from focus groups with members of the target population. Although this number may be an underestimate, as some articles might not specify that they used focus groups in the abstract, it does suggest that their use is relatively rare. One can certainly imagine that many more than 23 new instruments were introduced in the literature during this time period.

Moreover, this literature review identified only one article that described how focus groups can be applied in scale construction (O'Brien, 1993a; but see also Powell, Single, & Lloyd, 1996, who presented focus group findings and argued for their applicability to instrument development). Although certainly a step in the right direction, this article did not provide a detailed discussion of the benefits and limitations of focus groups or describe how data acquired from focus groups can be analyzed and implemented in item development. Thus, there is clearly a need for an in-depth examination of the potential of focus groups to inform instrument development.

Focus Groups

As mentioned previously, the focus group is a technique that involves a moderator-facilitated discussion among multiple participants about a specified topic of interest. Focus groups generate qualitative data that can be used to both enrich and extend what is known about a concept and inform item development. In turn, this knowledge can improve the relevance and representativeness of items.

Strengths and limitations of focus groups. Just as cultural anthropologists draw on ethnographic or emic methods that involve talking with members of the population under study to get an "insider" perspective about the lives and experiences of different cultures (Pike, 1954), focus groups provide a methodology that can allow researchers to learn about the meaning of a construct from the perspective of the population under study. This is important because researchers may unconsciously interpret the experiences of other "cultures" through the lens of their own cultural beliefs and values, leading to ethnocentric assumptions about the meaning of other cultures' behaviors and experiences that are inaccurate and incomplete (Podeolefsky & Brown, 1999). Presumably, the use of focus groups can allow researchers to discover the units of conceptualization from the perspective of the people studied, rather than imposing them from different cultural classifications of behavior (Pelto, 1970). This has been described as a "phenomenological approach," in which the goal is to obtain an understanding of the phenomenon as the respondents see it (Byers, Zeller, & Byers, 2002; Calder, 1977).

Although most of the support to date for the efficacy of focus groups is anecdotal rather than empirical, advocates have suggested that focus groups offer many potential advantages in comparison to one-to-one interviews. Among these advantages is the fact that participants not only respond to questions posed by the

moderator but they also respond to the comments of other focus group participants. They both query and explain themselves to each other. This makes focus groups more than simply the sum of separate individual interviews (Morgan, 1996). Because of the interaction among focus group members, a more in-depth discussion of events may be elicited than that obtained in a one-to-one interview, in which a sole participant answers the questions of an interviewer (Basch, 1987; Gray-Vickrey, 1993; Greenbaum, 1998). In addition, because participants may feel more comfortable discussing experiences with similar others, a more open and honest discussion might ensue than in a one-to-one interview with a researcher who is not a member of the target population (Basch, 1987). Moreover, several researchers have suggested that group pressures may inhibit individuals from providing misleading information (Basch, 1987; Millward, 2000), and focus groups may allow the investigator to learn about contrasting viewpoints in the population as group members react to each other's comments (Morgan, 1988; O'Brien, 1993a). Focus groups can also be both more cost effective and time effective than individual interviews, as multiple members of the population can be consulted at one time (Basch, 1987).

In turn, the information obtained from focus groups can be used to elaborate conceptualizations of key constructs and identify content that can be incorporated in item development (O'Brien, 1993a). In addition, knowledge gained about the language that members of particular populations use to describe their experiences (Millward, 2000; Podolefsky & Brown, 1999; Tilden, Nelson, & May, 1990) can inform the phrasing of items (O'Brien, 1993a). In one-to-one interviews participants may arguably adopt the language of the researcher.

An additional point regarding the usefulness of focus groups bears mentioning. Focus groups may be particularly beneficial for the development of measures of composite variables. In contrast with hypothetical constructs, in which a "postulated attribute of people" (Cronbach & Meehl, 1955, p. 283) is presumed to be responsible for observable behaviors, composite variables are represented by discrete, possibly uncorrelated, experiences that together "cause" or define the construct (Bollen & Lennox, 1991; MacCallum & Browne, 1993; Netland, 2001; Reckase, 1996). For example, with "intelligence," we assume that observed behaviors (effect indicators such as scores on a measure of intelligence or self-ratings of intelligence) are caused by the underlying construct. "Intelligence" is a construct that explains observed covariation among its indicators. Conversely, "life stressors" may be considered a composite variable, such that causal indicators (e.g., exposure to combat, experience of divorce) represent discrete stressor events that do not necessarily covary but that together compose the construct.

For the purpose of assessing a hypothetical construct such as "intelligence," the goal is to develop items that are equally good examples of identified content domains. As long as the content domain is adequately covered, the omission of one particular sample from this domain is not critical and may even be desirable from a practical perspective. Given equal content coverage, an instrument with fewer items is certainly preferable to one with many items. On the other hand, each item developed to assess a composite variable such as "life stressors" is intended to represent a discrete stressor event that is critical to the underlying construct. Omitting a causal indicator is equivalent to omitting a part of the construct (Bollen & Lennox, 1991). As discussed previously, focus groups allow for a rich discussion among group members in which important aspects of constructs are likely to emerge. Thus, one way that researchers may gain confidence that they have not missed critical aspects of composite variables is to implement focus groups in their content validation procedures.

Despite these strengths, there are several potential limitations associated with the focus group methodology. For one thing, some topics or populations may not lend themselves to focus group discussions.[1] As an example, members of certain psychopathological populations (e.g., individuals with avoidant personality disorder, social phobia, panic with agoraphobia, schizophrenia) might not be good candidates for focus group discussions about their disorders. Focus groups with members of the target population are probably also less useful for the discussion of very rare constructs such as certain types of phobias or unusual behaviors or lifestyles. Under these circumstances, recruiting participants may pose a major challenge. In addition, focus groups may not always be appropriate for topics considered too personal to share among strangers, such as sexual activities or behavior that may be considered socially deviant, such as drug use or sexual misconduct (Farquhar, 1999). Relatedly, focus groups may be more useful for "telling it like it is," while interviews are perhaps better for "telling it like it feels" (Michell, 1999; for a different perspective, see Byers et al., 2002; Carey, 1994; O'Brien, 1993b).

To the extent that a topic is sensitive, focus group participants' comfort and candidness may depend, at least in part, on the person who is asking the questions. In general, findings suggest that both female and male participants tend to prefer female interviewers, regardless of whether that individual is a member of the population herself (Farquhar, 1999; Spencer, Faulkner, & Keegan, 1988). Yet, one can imagine some topics for which a male moderator might be more appropriate. For instance, a study examining men's attitudes toward women might benefit from the use of a male moderator. Focus group participants' comfort and candidness may also be influenced by whether the participants have ongoing social contact outside the research context (Farquhar, 1999). Participants may feel more comfortable disclosing certain types of information to anonymous others than to those which whom they will continue to have contact.

Other factors may also limit the usefulness of data obtained from focus groups. For example, a moderator who asks questions in a leading manner can bias findings. The inclusion of an extremely dominant participant can produce results that are not reflective of the entire groups' experiences and perspectives. Group interaction may produce conformity pressures, distorting the individual's genuine perception of events (Byers et al., 2002) or simply limiting the information that participants are willing to provide. In addition, the nature of focus groups (bringing people together who share certain experiences and/or characteristics) may result in a discussion that extends beyond the target construct. There are several things the researcher can do to minimize these potential problems. The researcher can select a moderator who is not only intimately familiar with the goals of the research study but who is also knowledgeable regarding the need to take an objective, yet empathic, stance (Krueger, 1998a). Dominant group members can be "managed" by skillful moderators. Conformity effects can

---

[1] We thank an anonymous reviewer for suggesting several circumstances in which focus groups might be less helpful.

be reduced by asking participants to specify how their views differ from the perspectives expressed by others. A strong moderator who is well acquainted with the guiding conceptualization of key constructs can refocus the conversation when it gets off track (although, as is detailed in our presentation of data analytic issues, the researcher may be best served by allowing some discussion of topics that appear to be peripheral to the target construct at this stage). Of course, even in the best circumstances, some focus groups may be more productive than others, and researchers are encouraged to seek corroboration for information obtained from focus groups, just as they seek corroboration for other types of data they collect.

Another potential weakness of focus groups pertains to the representativeness of findings. Because the number of focus groups one can conduct is necessarily limited, one may question the generalizability of the information obtained in focus groups. Thus, it is important to clarify that the goal of focus groups is not, and should never be, to represent the views of a population. For the purpose of instrument development, the goal is to generate information that can be used to expand what is known about a construct and to inform item development. Also, given the qualitative nature of focus group data, many researchers are concerned that data analysis may be unduly influenced by the researcher's subjective views. Researchers may not be aware that scientific data analysis techniques are available. Below, we review common methods of focus group analysis and present a method that may be particularly useful for the application of focus groups in instrument development.

Analyzing focus group data. Although methods for analyzing other forms of qualitative data (e.g., depth interviews and ethnographic field notes) are well developed, methods for analyzing focus groups that are practical, systematic, and verifiable have only recently become available (Frankland & Bloor, 1999; Krueger, 1994). Yet, to our knowledge, no one has yet elaborated a systematic analytical technique specific to the use of focus groups in instrument development. This is unfortunate, given that researchers have long encouraged the application of multiple methods in research (i.e., triangulation), and several researchers have noted that focus groups may be helpful in constructing questionnaires (Barbour & Kitzinger, 1999; O'Brien, 1993a; Powell et al., 1996).

Just as for quantitative analysis, there are many possible approaches to focus group data analysis, and the approach one chooses depends on the purpose of the study (Krueger, 1994, 1998b). Although some researchers may desire to use focus group data in a manner that translates these data into numbers and some experts endorse this approach (e.g., Millward, 2000), many focus group proponents prefer to maintain the qualitative nature of these data, arguing that numbers can be misleading in focus group analysis, in which the goal is to obtain theoretical saturation and not generalizability (Asbury, 1995; Barbour & Kitzinger, 1999; Byers et al., 2002; Calder, 1977; Krueger, 1998b).

There are three main questions a researcher must ask in conducting focus group analysis. The first question regards how the data from the focus group will be acquired. The researcher may choose between four options: transcript-based analysis, tape-based analysis, note-based analysis, and memory-based analysis (Krueger, 1998b). For the purpose of instrument development, we recommend that the focus group data take the form of a tape-based abridged transcript. In other words, we suggest that researchers

audiotape each focus group discussion and create a written record of the relevant and useful portions of the discussion (on the basis of themes specific to the focal constructs). Videotaping is certainly possible and might be preferred because it is a more comprehensive source of data; however, some researchers have suggested that videotaping may be particularly inhibiting to focus group members (Barbour & Kitzinger, 1999).

The second question relates to the level at which the results will be reported. The researcher can report some combination of raw data, description, interpretation, and recommendations (Krueger, 1998b). For the purpose of scale construction, we recommend that data be reported at a descriptive level, in which themes are presented and illustrative quotes from the abridged transcript are provided. These data can then inform the conceptualization of key constructs and item development. One challenge researchers face at this stage is determining which information is relevant and useful. As discussed previously, bringing together a group of individuals who share similar background characteristics may result in the acquisition of some information that is not directly relevant to the target construct. On the one hand, a researcher who is too influenced by new information may err on the side of adopting a conceptualization that is overly influenced by a particular group or groups of individuals and that may ultimately include facets that are only spuriously related to the target construct. On the other hand, a researcher who too rigidly attends to the a priori conceptualization runs the risk of providing an overly conservative parameterization of the construct and of missing out on newly identified facets of key constructs. At this initial stage in content validation, we recommend a liberal approach to the inclusion of additional content domains, and we caution against narrowing conceptualizations on the basis of information obtained from focus groups. During later psychometric evaluation (e.g., when examining quantitative item characteristics), items sampled from peripheral content domains may likely be eliminated.

The final choice concerns whether the data are to be analyzed by hand or by computer (Krueger, 1998b). Popular qualitative data analysis packages, such as The Ethnograph (Soclari, 2003a) or QSR N6 (Soclari, 2003b) are helpful because they ensure that the analysis is based on the entire universe of relevant textual material (Frankland & Bloor, 1999). For a small amount of transcript material (as might be the case for fewer than six focus groups), simple manual indexing may be an economical alternative to a qualitative data analysis package (Frankland & Bloor, 1999). In fact, as is detailed shortly, this is the procedure we chose for our demonstration of how focus groups can be integrated in scale construction. Although the context for this example involves research on war-related stressors, this procedure may be easily adapted by researchers for other purposes. For a more detailed description of focus group analysis options, please see Krueger (1998b).

Application of Focus Groups in Instrument Development: An Example

Background: Deployment Stress Studies

Our goal for the project that serves as a vehicle for the presentation of how focus groups can be used in content validation was to develop an instrument to assess a broad spectrum of war-related stressors that would be content valid for veterans of contemporary

military deployments. As mentioned previously, there is some evidence for the limited content validity of instruments that have been used for the investigation of war-related stressors. More specifically, several researchers have criticized the literature on Vietnam veterans because of the nearly exclusive use of measures of combat exposure to index war-related stress exposure (Kaylor, King, & King, 1987; D. W. King & King, 1991; D. W. King, King, Gudanowski, & Vreven, 1995). Researchers have pointed to other potentially important war-related stressors, such as the persistent and pervasive fear of bodily harm or death, exposure to the devastation to life and property observed in the aftermath of battle, and the day-to-day discomforts that typically accompany combat duty (Breslau & Davis, 1987; Gallers, Foy, Donahoe, & Goldfarb, 1988; Hendin & Haas, 1984; D. W. King et al., 1995; Paul, 1985; Stretch, Vail, & Mahoney, 1985).

Despite the call to broaden the assessment of war-related stressors, much of the more recent research on veterans of the 1990–1991 Gulf War (Gulf War I) has also focused on combat exposure, with the majority of these studies relying on measures originally developed for use with Vietnam veterans (Marshall et al., 1999). Most prominently, researchers have used Keane et al.'s (1989) Combat Exposure Scale. This seven-item instrument is aimed at assessing traditional combat experiences and includes items such as "How often did you fire rounds at the enemy?" and "What percentage of the men in your unit were killed, wounded, or missing in action?" Although this measure has demonstrated validity with Vietnam veterans, it is less appropriate for the assessment of the combat experiences of Gulf War I veterans. The nature of warfare has changed dramatically since the Vietnam War, and as such, the combat experiences of veterans of more recent deployments, such as Gulf War I, likely differed markedly from those of Vietnam veterans (Marshall et al., 1999; Norwood & Ursano, 1996). For example, the possibility of nuclear, biological, and chemical exposures, and fear associated with possible exposures, is a growing concern for contemporary deployments. Moreover, because of advances in technology, many veterans of more contemporary cohorts may not experience the same level of direct combat as previous cohorts. In addition, the deployment of a much larger proportion of women and National Guard and Reservist forces in more recent deployments calls attention to stressors such as sexual harassment and concerns about life and family disruptions that were not considered of research importance for previous cohorts of veterans. Thus, not only may the items in the Combat Exposure Scale not discriminate at an appropriate level of intensity (Haynes et al., 1995; Smith & McCarthy, 1995) for more recent veteran cohorts but a broader assessment of war-related stressors is needed to better understand the impact of deployment on veterans' health and well-being.

Overview of Content Validation Procedure

The process of content validation that we used in our development of measures of war-related stressors involved four primary components: a review of relevant literature, the ongoing refinement of definitions of key constructs, the conduct of focus groups with members of the target population, and the development of items for subsequent administration in survey form. It is important to note that activities were for the most part simultaneous and iterative rather than strictly linear. The process of developing construct definitions both preceded and followed from the focus groups. Likewise, a preliminary item pool proposed prior to the conduct of focus groups was refined and enhanced following focus groups and elaboration of construct definitions. In addition, the literature was consulted early on to identify preliminary conceptualizations of key constructs (Clark & Watson, 1995) and to identify relevant item sets. It was consulted later to provide corroboration for the findings of the focus groups. Below, we provide a detailed description of the focus group component of this procedure.

Focus Groups

Prior to conducting focus groups, and on the basis of our review of the literature, we identified six constructs that we felt provided a preliminary conceptual foundation: exposure to stereotypical combat events, exposure to the aftermath of battle, perceived threat, difficult living and working environment, concerns about life and family disruptions, and sexual harassment (please refer to Appendix A for full definitions for these and other stressor constructs). Our goal was to assess both objective stressors, such as combat and aftermath of battle, and subjective stressors, such as concerns about life and family disruptions and perceived threat (Green, 1994; D. W. King et al., 1995; Ozer, Best, Lipsey, & Weiss, 2003; Solomon, Mikulincer, & Hobfoll, 1987). In addition, we did not restrict our conceptualization to factors specific to being in a war zone (e.g., combat, aftermath of battle) but were interested in any psychosocial stressor that may have been significant for military personnel over the course of their deployment (e.g., concerns about life and family disruptions, sexual harassment). After we had developed preliminary definitions of these war-related stressors, we conducted focus groups with Gulf War I veterans. Focus group participants were volunteers recruited from computerized registries of veterans who had used Veterans Administration facilities for medical care. Generally, 8–10 members per focus group are recommended to ensure adequate discussion among group members (Greenbaum, 1998), but even fewer focus group participants may be appropriate for topics for which more in-depth information is needed (Barbour & Kitzinger, 1999). We were able to enroll an average of approximately 7 individuals per group, a number that we found sufficient to foster a rich discussion among group members.

As recommended (Morgan, 1997), we conducted focus groups in a relatively informal and comfortable setting with focus group members seated around a circular table to facilitate the participation of all members of the group. It is important to use a focus group moderator who is well-informed of the goals of the research project, skilled in moving the discussion through the major topic areas, and able to engage all participants (Krueger, 1998a). On the basis of these criteria, we selected a moderator who was a member of our research team. In order to encourage a diversity of opinions among group members, prior to beginning the focus groups we asked participants to speak up if another member of the group expressed an opinion that did not represent their own experience and reassured participants that there were no right or wrong answers. During the focus groups, the moderator first asked general questions to allow participants to guide the conversation, such as "What is the first thing that comes to mind about your experience in the Gulf War?" Themes (i.e., a priori war-related stressors gleaned from the existing literature) were probed to help guide the discussion only when participants did not spontaneously generate

them. For example, to probe concerns about life and family disruptions caused by deployment to the Gulf region, the moderator asked participants, "What was it like to be away from your family and life back home while you were in the Gulf War?" The moderator broached the issue of sexual harassment by first noting that "In the past couple of years there has been a lot of media attention to sexual harassment in the military," after which participants were asked "Do you think that this was a problem in the Gulf War?" The focus group guide evolved over the course of the focus groups, and questions were included to probe additional constructs as they were identified. The full guide that was used for the final focus group with male personnel deployed from the Reserves and National Guard is included in Appendix B. Focus groups were approximately 1.5–2.0 hr in duration. Participants were informed prior to the beginning of the discussion that the group was intended for the purpose of collecting information but was not meant to be a therapy session. Of course, we encouraged participants to let us know if they were disturbed by the content of the discussion so that we might provide a referral to a trained clinician if necessary. None of our participants indicated the need for a referral.

Typically, three to five groups are needed to ensure coverage of the range of experiences or opinions about a particular topic (Morgan, 1997). We conducted six focus groups in total. After conducting one focus group that was mixed by predeployment duty status (i.e., deployment from active duty versus activation from the National Guard or Reserves) and gender, and as recommended by focus group experts (Morgan, 1996), we constituted the remaining focus groups to include participants who were more homogenous with regard to these key background characteristics to foster a richer discussion among group members. In addition, we conducted focus groups with members of several key subgroups to ensure that important variations in the population were reflected in the sample (Byers et al., 2002). More specifically, we conducted three groups with male veterans who were deployed to the Gulf region from active duty units, one group of male veterans who were activated from the National Guard or Reserves, and one group of female Gulf War I veterans. We reached a saturation point after our fifth group; the same issues and topics began to emerge, providing some evidence for the reliability of our focus group data (Byers et al., 2002). At this point, we determined that the amount of new information obtained from the sixth group was less valuable than the time and effort that was required to obtain the information. Thus, we felt confident that we had adequately covered key content domains at this point, and completed the focus groups.

To facilitate the coding and analysis of data, all sessions were audiotaped with the consent of participants. We used a tape-based analytic strategy that involved developing an abridged transcript of the relevant and useful portions of the discussion (Krueger, 1998b). The relevance of the discussion was determined by whether it contained a reference to any of our previously identified war-related stressor themes or by whether it introduced an additional war-related stressor that we had not previously considered. To clarify: Three individuals, two of whom had been physically present during the focus groups, listened to an audiotape of each focus group. This approach is consistent with the recommendation that at least one person who was physically present in the room when the focus group was conducted and who is familiar with the context of the discussion participate in data analysis (Krueger,

1994). Coders were provided with a list of themes (i.e., war-related stressors) and definitions that they were told may or may not have been discussed during the focus group. Coders listened to audiotapes twice. First, they simply listened to become familiar with the flow of the conversation and the topics that were raised. On the second review, coders were asked to stop the audiotape each time they heard a participant mention an identified war-related stressor and to record verbatim the sentence or sentences in which it was discussed. Coders were also instructed to stop the audiotape if they heard a participant describe any other war-related stressor that was not included among our previously identified list and to record this sentence or sentences as well. Finally, coders noted any novel terminology participants used to describe their experiences.

Next, the coders met to discuss their review of the tapes. There was a great deal of consensus among coders regarding the quotes that were examples of identified constructs, providing additional evidence for the reliability of the obtained data (Krueger, 1998b). To the extent that coders agreed with one another regarding whether quotes were examples of identified themes (i.e., constructs), quotes relevant to each of the key constructs were compiled. Coders also introduced additional constructs for consideration in these meetings. As recommended by focus group experts (Krueger, 1994), newly nominated stressors and ongoing refinements of conceptualizations were incorporated in the list of war-related stressors and definitions used by coders throughout the process. The information in the final compilations was then used to refine definitions of constructs as needed, to identify additional constructs, and to inform item development. Examples of how this information was used are provided below.

Construct refinements and additions based on focus groups. The information gleaned from focus groups provided corroboration for the constructs that we had initially identified from our literature review. In addition, we were able to considerably elaborate and refine our preliminary definitions of war-related stressors after reviewing the information generated from focus groups. Thus, the process of construct definition was iterative. Particularly with novel constructs such as those that were the focus of this study, initial working definitions may require considerable revision over the course of the content validation process. Focus groups were particularly useful in refining four of our conceptualizations: combat, perceived threat, aftermath of battle, and difficult living and working environment. *Combat* was initially defined as exposure to hostile acts of war perpetrated by the enemy but was revised to include experiences of "friendly fire" incidents, the only form of combat that many of our participants reported. Relatedly, the recognition of the threat of friendly fire incidents informed our definition of perceived threat. After hearing quotes like "We got into a crossfire with ourselves because we couldn't get everyone to shut down at the same time," we realized the need to include perceived threat of incidents of friendly fire in the definition of this construct.

We also refined our conceptualization of *aftermath of battle*, initially defined as exposure to the consequences of combat and including experiences such as seeing dead or injured soldiers and handling human remains. Many focus group participants, however, mentioned distressing contact with prisoners of war in the aftermath of battle. For instance, one participant said "We had wire, big circle with 600 prisoners, they could have jumped over if they wanted to leave—it was sad—those people didn't want to fight." We thus expanded our definition of *aftermath of battle* to include

experiences with prisoners of war. In addition, our definition of difficult living and working environment was refined to include facets of this construct that emerged through focus groups: for example, cultural difficulties, lack of privacy, and inadequate equipment.

Using information generated from focus groups (and with on-going attention to the literature), we were able to identify several additional war-related stressors. For example, our focus group with female veterans revealed another stressor specific to the experience of being a woman in the military. Although we had previously identified sexual harassment as a significant stressor, we had not considered that women might have been exposed to harassment that was associated with their biological sex or social status in the military but that was not sexual in nature. Yet, several female veterans reported experiencing harassment of this kind. One participant told a story about being harassed by her sergeant.

> Well, there was this sergeant who definitely didn't like women in the military in general, but he was my platoon sergeant—while the guys were lounging around in their tents, he always used to give us duties, "Go take the tires apart, or go do this, or go do that, go dig the bunkers." Our assistant platoon sergeant was the same way.

Although this construct emerged in the focus group with women, we thought that it might also represent a stressor for some men. Thus, we developed the construct of general harassment to address exposure to harassment that is nonsexual in nature but that may occur on the basis of one's biological sex or social status and that is used to enforce traditional social roles or in response to the violation of these roles.

The conceptualization of constructs was iterative within this study. Although researchers have noted the iterative nature of construct definition over the course of many studies (Haynes et al. 1995), we further note that initial working definitions, particularly those of novel constructs, may require considerable revision during the instrument development phase of a single study. The iterative item construction process was repeated for each of the war-related stressors and extended over a 1-year period of time (although it certainly would have been feasible to complete this process in a shorter period of time).

Item development. Information from focus groups was also used to inform item development via the application of a rational approach to test construction (Jackson, 1971; Nunnally & Bernstein, 1994). This information was first used to affirm items drawn from existing measures of related constructs. Then, guided by our elaborated definitions and with ongoing reference to the literature, this information was used to refine and supplement these sets of candidate items. Specifically, relevant quotes were reviewed to identify content that could be incorporated in item statements. For instance, the quote

> I don't know if you can ever get used to eating watered food, you couldn't even cook it because they wouldn't let you light any fires—so you're eating dehydrated beef patties from Vietnam that are starting to taste good after a while

was used to develop the item "The food you had to eat was of very poor quality (for example, bad or old MREs)." The quote

> I had a 4-year-old son at the time who lost his father—it was very hard—missing him losing his teeth, you know, just the little things, Christmas, and not being able to talk on the phone

was used to develop two items: "While I was deployed, I was concerned about missing out on my children's growth and development" and "While I was deployed, I was concerned about missing important events at home, such as birthdays, weddings, funerals, graduations, etc." The quote

> The biggest deal with men in the Middle East was that women should not be allowed to leave the compound without pants on (i.e., in shorts). . . . I'm like "Hello! We're in a war!"

was used to develop the item "My daily activities were restricted because of local religious or ethnic customs." The quote

> It's [anticipation's] one of the biggest things I feel like I have to explain—at the time we were told they're the third largest enemy in the world, they've been sitting here for 6 months, they're dug in, they're waiting for you

was used to develop the item "I felt that I was in great danger of being killed or wounded."

It is important to note that consistent with the experience reported by other researchers who have used focus groups to generate content for instrument development (O'Brien, 1993a; Powell et al., 1996), many of the items we developed contained specific details that were not readily available in the literature, such as details regarding family concerns and aspects of the difficult living and working environment. Special attention was given to the inclusion of language and colloquialisms used by veterans who participated in our focus groups. For example, several focus group participants referred to nuclear, biological, and chemical agents as "NBCs," and we integrated this acronym into our items (along with the full spelling of the term for those who might be unfamiliar with the acronym). We attempted to sample all possible content domains within each construct, erring on the side of overinclusiveness, with the intention of eliminating items that might prove weak or tangentially related to the construct of interest at a later point in the psychometric evaluation of the instrument (Clark & Watson, 1995). If we felt that a quote provided information that was not addressed by a preexisting item, we developed an item. Once we had supplemental item sets for each construct, we eliminated redundant items and examined the item pool to ensure that there was a proportional representation of facets within constructs and thus that items could be aggregated to provide an adequate assessment of the underlying construct. Although we did not develop specific decision rules for inclusion or exclusion of information gained from focus groups, this is certainly an option that others researchers might consider in their own work.

As recommended (Lynn, 1986), items were provided to content-validity experts and members of the target population (i.e., Gulf War I veterans), who were asked to individually review the items for their clarity and relevance to the target constructs. Of course, the ideal-case scenario, albeit one that might be difficult to obtain, would be to have an item review by members of the target population who possess expertise in the topic area (Kubany, Abueg, Kilauano, Manke, & Kaplan, 1997). In fact, one of our reviewers met this criterion. In line with recommended procedures (Clark & Watson, 1995; Haynes et al., 1995), the item review was iterative in nature. In other words, content-validity experts had the opportunity to reexamine revised items. Often, items are subjected to review, revised, and never subjected to review again (Clark & Watson, 1995), which can limit content validity, as revisions may fix one problem but introduce yet another.

Psychometric Characteristics of Measures and Incremental Validity of Approach

Clearly, conducting focus groups can be a time-consuming and costly endeavor (although, as stated previously, they are certainly more time- and cost-effective than individual interviews). Thus, one must question whether the benefits of using focus groups to inform instrument development outweigh the costs of conducting them (Yates & Taub, 2003). One concern is whether measures that draw on information derived from focus groups demonstrate acceptable psychometric characteristics. With regard to the war-related stressor measures that are here provided as an example of how focus groups can be applied in instrument development, we examined initial psychometric properties of preliminary war-related stressor item sets in a test development sample of approximately 350 Gulf War I veterans. We refined the measures by deleting items with weaker characteristics (i.e., typically those for which item-total correlations were low or for which there was little dispersion), as appropriate. Scale characteristics for the refined measures were quite good, given the relative brevity of each measure (total number of items ranged from 7 to 20); internal consistency reliability estimates ranged from .82 to .91.

Evidence for criterion-related validity was derived from a second validation sample of 357 Gulf War I veterans. Findings revealed modest to moderate associations between war-related stressors and a collection of health outcomes (e.g., average correlation for satisfaction with life, $r = -.25$; depression, $r = .24$; anxiety, $r = .27$; posttraumatic stress disorder, $r = .33$; and neurocognitive deficits, $r = .26$). Thus, these scales have shown quite acceptable internal consistency reliability, as well as satisfactory criterion-related validity vis-à-vis the prediction of mental and physical health outcomes (L. A. King, King, Vogt, Knight, & Samper, 2004).[2]

One might also question the extent to which the introduction of focus groups in instrument development provides a meaningful improvement over and above typical development procedures (i.e., the incremental validity of focus groups; Hunsley, 2003; Sechrest, 1963). In the context of the war-related stressor measures, *incremental content validity* refers to the degree to which the elements of the assessment instruments (e.g., items) tap a broader range of facets of target constructs as a consequence of the focus groups (Haynes & Lench, 2003). As described previously, focus groups were useful in identifying one additional war-related stressor. Moreover, we identified additional facets for four of the six constructs that composed our preliminary conceptual foundation. In turn, new items were developed to assess these content domains. One new item was developed to supplement the initial pool of items for combat; six new items were developed to assess aftermath of battle; nine new items were developed to assess perceived threat; and seven new items were developed to assess difficult living and working environment.

Another form of incremental validity, *incremental criterion validity*, refers to the degree to which a newly developed measure accounts for a higher proportion of variance in a criterion than existing measures (Haynes & Lench, 2003). In the case of our war-related stressor measures, this refers to the extent to which measures that include items derived from focus groups demonstrate higher criterion validity than they did before adding new items. Analyses of incremental criterion validity ordinarily involve hierarchical multiple regressions to document the contribution of newly developed measures to the prediction of outcomes above

and beyond what is accounted for by existing measures. Because our goal was to improve existing item sets, we computed correlations for each of our scales both before and after adding items. Thus, we computed indices of incremental criterion validity that represented the average correlation coefficient between the original and revised war-related stressor measures and a number of health outcomes (i.e., satisfaction with life, depression, anxiety, posttraumatic stress disorder, and neurocognitive deficits). For combat, four of the five validity coefficients increased from the original to the revised measure (average validity coefficients increased from .25 to .26). For aftermath of battle, two of the five validity coefficients increased (average validity coefficients were .43 for the original and revised item sets). For perceived threat, four of the five validity coefficients increased (average validity coefficients increased from .39 to .46). For difficult living and working environment, all five validity coefficients increased (average validity coefficients increased from .39 to .44). It is interesting to note that these changes were accompanied by only a slight increase in internal consistency reliability for aftermath of battle and difficult living and working environment and no change for combat.[3] We argue that the stability of internal consistency reliability accompanied by an increase in criterion validity provides some evidence for incremental criterion validity.[4] That is, the improvement in validity cannot solely be attributed to an increase in reliability for these three scales. Instead, the enhanced validity of these scales may likely be attributed to the inclusion of focus groups and the more complete sampling of content in the content validation procedure. Despite the post hoc nature of these analyses (i.e., this project was not designed for the purpose of examining incremental validity), we think that they provide encouraging support for the incremental validity of adding focus groups to the instrument development procedure.

---

[2] The measures and a manual (D. W. King, King, & Vogt, 2003) describing their development and psychometric properties are available upon request from Dawne S. Vogt.

[3] For combat, the internal consistency reliability coefficient was .85 for both the old and new item sets. For aftermath of battle, perceived threat, and difficult living and working environment, internal consistency reliability coefficients increased from .85 to .89, .56 to .85, and .87 to .89, respectively. The substantial improvement in internal consistency reliability for perceived threat may likely be a consequence of the dramatic increase in the number of items, from 6 items before focus groups to 15 items at the conclusion of the instrument development procedure.

[4] As additional support, we projected the validity coefficients for the original item sets to values that would obtain had the reliability of the original item sets been that of the revised item sets, and we then compared these projected validity coefficients with the validity coefficients calculated for the revised item sets. Support for incremental validity would be evidenced if the projected validity coefficients based on the original item sets were less than the actual validity coefficients using the revised item sets. Indeed, of the 20 validity coefficients (5 for each of 4 stressor measures), 12 were smaller, 3 were the same (within rounding), and 5 were larger. Assuming an expected equal likelihood for each of these three possibilities, the observed frequencies yielded $\chi^2(2, N = 20) = 6.79$, $p < .05$. Thus, the increase in validity coefficients for the revised item sets vis-à-vis the projected validity coefficients is greater than that which can be attributed to enhanced reliability simply as a function of increases in the number of items.

Summary and Recommendations for Future Research

In summary, we advocate for communication with, and information gathering from, members of the target population in instrument development. We especially recommend focus groups to obtain information that can be used to inform the definition of key constructs, the identification of additional constructs and elaboration of content domains, and translations into appropriately worded item statements. This information can also be used to corroborate conceptualizations of key constructs and items from existing measures. Although some researchers have incorporated consultation with members of the target population in their content validation procedures, most fail to bring consultation in early enough to inform the identification and specification of key constructs. We suggest that consultation at an earlier stage can contribute to the development and use of measures that have optimal content validity for their intended purpose.

Of course, we are not suggesting that focus groups are the only avenue through which one can consult members of the target population. Instead, we argue that focus groups can provide a nice alternative to one-to-one interviews in the content validation of newly developed instruments. To reiterate a point made earlier: Focus groups may be more appropriate for some purposes and less appropriate for other purposes, and the choice to use them ultimately depends on the researcher's needs and preferences. Also, although not addressed in the present paper, focus groups can provide useful information for theory building and hypothesis development (O'Brien, 1993a).

Whatever procedure one ultimately adopts, it is imperative that researchers who present new instruments in the literature provide a detailed account of their content validation procedure and identify the functions and populations for which any newly developed instrument has validity, as well as the limits of content validity. As Haynes et al. (1995) noted, the content validity of a particular assessment device is always conditional. A measure that is valid for one population may not be valid for another population, and a measure that is valid for one purpose may not be valid for another purpose. Our review of the articles published in *Psychological Assessment* from 1995 to 2002 revealed that although there has been some improvement in the documentation of content validation procedures, ambiguity surrounding the procedures used by researchers remains.

On a related note, authors of newly developed measures must carefully consider the level of specificity at which they define their constructs and the implications that this decision has for content validity. Although it is often difficult to determine the best level of specificity, a functional approach to assessment suggests that the level of specificity depends on the goals of the assessment (Haynes & O'Brien, 2000). For the example in this article, our goal was to develop scales to assess war-related stressors that would be content valid for contemporary deployments. At a higher level of specificity, we might have attempted to develop measures that assessed, for example, war-related stressors specific to women. Had we used an appropriate content validation procedure, the result would have been a measure that was highly content valid for women but that was likely less useful for assessing the war-related stressors experienced by men (Haynes & O'Brien, 2000). Similarly, at a lower level of specificity, we might have attempted to develop a measure of war-related stressors common to all past and present U.S. military deployments. However, this approach probably would

have resulted in a measure unable to capture the unique characteristics of more contemporary deployments. Thus, we chose a level of specificity that fell between these two extremes. Of course, the content validity of this instrument for future military deployments will need to be assessed, and it can be expected that content validity will diminish as deployment-related stressors evolve over time.

We conclude by recapitulating a point made earlier: Content validity is but one step—albeit an important one—in the process of construct validation. Measures that are content valid for their intended purpose and that possess other requisite psychometric properties have the potential to make a substantial contribution to the literature.

References

Asbury, J. E. (1995). Overview of focus group research. *Qualitative Health Research, 5,* 414–420.

Barbour, R. S., & Kitzinger, J. (1999). *Developing focus group research.* Thousand Oaks, CA: Sage.

Basch, C. E. (1987). Focus group interview: An underutilized research technique for improving theory and practice in health education. *Health Education Quarterly, 14,* 411–448.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equations perspective. *Psychological Bulletin, 110,* 305–314.

Breslau, N., & Davis, G. C. (1987). Posttraumatic stress disorder: The etiological specificity of wartime stressors. *American Journal of Psychiatry, 144,* 578–583.

Byers, P. Y., Zeller, R. A., & Byers, B. D. (2002). Focus group method. In M. W. Wiederman & B. E. Whitley, Jr. (Eds.), *Handbook for conducting research on human sexuality* (pp. 173–193). Mahwah, NJ: Erlbaum.

Calder, B. J. (1977). Focus groups and the nature of qualitative marketing research. *Journal of Marketing Research, 14,* 353–364.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist, 15,* 546–553.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Carey, M. A. (1994). The group effect in focus groups: Planning, implementing and interpreting focus group research. In J. M. Morse (Ed.), *Critical issues in qualitative research methods* (pp. 225–241). London: Sage.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

DeVellis, R. F. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Fagot, B. I. (1992). Assessment of coercive parent discipline. *Behavioral Assessment, 14,* 387–406.

Farquhar, C. (1999). Are focus groups suitable for "sensitive" topics? In R. S. Barbour & J. Kitzinger (Eds.), *Developing focus group research* (pp. 47–63). Thousand Oaks, CA: Sage.

Frankland, J., & Bloor, M. (1999). Some issues arising in the systematic analysis of focus group materials. In R. S. Barbour & J. Kitzinger (Eds.), *Developing focus group research* (pp. 144–155). Thousand Oaks, CA: Sage.

Frank-Stromberg, M. (1989). Reaction to the Diagnosis of Cancer Questionnaire: Development and psychometric evaluation. *Nursing Research, 38,* 364–369.

Gallers, J., Foy, D. W., Donahoe, C. P., & Goldfarb, J. (1988). Posttraumatic stress disorder in Vietnam combat veterans: Effects of traumatic

violence exposure and military adjustment. *Journal of Traumatic Stress, 1,* 181–192.

Gray-Vickrey, P. (1993). Gerontological research: Use and application of focus groups. *Journal of Gerontological Nursing, 19*(5), 21–27.

Green, B. L. (1994). Psychological research in traumatic stress: An update. *Journal of Traumatic Stress, 7,* 341–362.

Greenbaum, T. L. (1998). *The handbook for focus group research* (2nd ed.). Thousand Oaks, CA: Sage.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456–466.

Haynes, S. N., Nelson, K., & Blaine, D. D. (1999). Psychometric issues in assessment research. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 125–154). New York: Wiley.

Haynes, S. N., & O'Brien, W. H. (2000). Specificity of variables. In S. N. Haynes & W. H. O'Brien (Eds.), *Principles and practice of behavioral assessment* (pp. 128–139). Norwell, MA: Kluwer Academic/Plenum Publishers.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7,* 238–247.

Hendin, H., & Haas, A. P. (1984). Combat adaptations of Vietnam veterans without posttraumatic stress disorders. *American Journal of Psychiatry, 141,* 956–959.

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15,* 443–445.

Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review, 78,* 229–248.

Kaylor, J., King, D. W., & King, L. A. (1987). Psychological effects of military service in Vietnam: A meta-analysis. *Psychological Bulletin, 102,* 257–271.

Keane, T. M., Fairbank, J. A., Caddell, J. M., Zimering, R. T., Taylor, K. L., & Mora, C. A. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment, 1,* 53–55.

King, D. W., & King, L. A. (1991). Validity issues in research on Vietnam veteran adjustment. *Psychological Bulletin, 109,* 107–124.

King, D. W., King, L. A., Gudanowski, D. M., & Vreven, D. L. (1995). Alternative representations of war zone stressors: Relationships to posttraumatic stress disorder in male and female Vietnam veterans. *Journal of Abnormal Psychology, 104,* 184–196.

King, D. W., King, L. A., & Vogt, D. S. (2003). *Manual for the Deployment Risk and Resilience Inventory (DRRI): A collection of measures for studying deployment-related experiences in military veterans.* Boston: National Center for PTSD.

King, L. A., & King, D. W. (1990). Role conflict and role ambiguity: A critical assessment of construct validity. *Psychological Bulletin, 107,* 48–64.

King, L. A., King, D. W., Vogt, D. S., Knight, J. A., & Samper, R. E. (2004). *Development and psychometric properties of the Deployment Risk and Resilience Inventory.* Manuscript submitted for publication.

Krueger, R. A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Beverly Hills, CA: Sage.

Krueger, R. A. (1998a). *Moderating focus groups.* Thousand Oaks, CA: Sage.

Krueger, R. A. (1998b). *Analyzing and reporting focus group results.* Thousand Oaks, CA: Sage.

Kubany, E. S., Abueg, F. R., Kilauano, W. L., Manke, F. P., & Kaplan, A. S. (1997). Development and validation of the sources of Trauma-Related Guilt Survey—War-Zone version. *Journal of Traumatic Stress, 10,* 235–258.

Kubany, E. S., Haynes, S. N., Abueg, F. R., Manke, F. P., Brennan, J. M., & Stahura, C. (1996). Development and validation of the Trauma-Related Guilt Inventory (TRGI). *Psychological Assessment, 8,* 428–444.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694.

Loo, C. M., Fairbank, J. A., Scurfieldl, R. M., Ruch, L. O., King, D. W., Adams, L. J., & Chemtob, C. M. (2001). Measuring exposure to racism: Development and validation of a Race-Related Stressor Scale (RRSS) for Asian American Vietnam veterans. *Psychological Assessment, 13,* 503–520.

Lynn, M. R. (1985). Reliability estimates: Use and disuse. *Nursing Research, 34,* 254–256.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35,* 382–385.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin, 114,* 533–541.

Malgady, R. G., Rogler, L. H., & Cortes, D. E. (1996). Cultural expression of psychiatric symptoms: Idioms of anger among Puerto Ricans. *Psychological Assessment, 3,* 265–268.

Marsella, A. J., & Kameoka, V. A. (1989). Ethnocultural issues in the assessment of psychopathology. In S. Wetzler (Ed.), *Measuring mental illness: Psychometric assessment for clinicians* (pp. 231–256). Washington, DC: American Psychiatric Association.

Marshall, G. N., Davis, L. M., & Sherbourne, C. D. (1999). *A review of the scientific literature as it pertains to Gulf War illnesses: Vol. 4: Stress.* Arlington, VA: RAND.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30,* 955–966.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14,* 5–8.

Michell, L. (1999). Combining focus groups and interviews: Telling how it is; telling how it feels. In R. S. Barbour & J. Kitzinger (Eds.), *Developing focus group research* (pp. 36–46). Thousand Oaks, CA: Sage.

Millward, L. J. (2000). Focus groups. In G. M. Breakwell, S. Hammond, & C. Fife-Schaw (Eds.), *Research methods in psychology* (2nd ed., pp. 303–324). London: Sage.

Mitchell, J. V., Jr. (1986). Measurement in the larger context: Critical current issues. *Professional Psychology: Research and Practice, 17,* 544–550.

Morgan, D. L. (1988). *Focus groups as qualitative research.* Newbury Park, CA: Sage.

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology, 22,* 129–152.

Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage.

Netland, M. (2001). Assessment of exposure to political violence and other potentially traumatizing events: A critical review. *Journal of Traumatic Stress, 14,* 311–326.

Norwood, A. E., & Ursano, R. J. (1996). The Gulf War. In R. J. Ursano & A. E. Norwood (Eds.), *Emotional aftermath of the Persian Gulf War: Veterans, families, communities, and nations* (pp. 3–21). Washington, DC: American Psychiatric Association.

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

O'Brien, K. (1993a). Improving Survey Questionnaires through focus groups. In D. L. Morgan (Ed.), *Successful focus groups: Advancing the state of the art* (pp. 105–117). Thousand Oaks, CA: Sage.

O'Brien, K. (1993b). Using focus groups to develop health surveys: An example from research on social relationships and AIDS-preventive behavior. *Health Education Quarterly, 20,* 361–372.

Ozer, E. J., Best, S. R., Lipsey, T. L., & Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychological Bulletin, 129,* 52–73.

Paul, E. A. (1985). Wounded healers: A summary of the Vietnam Nurse Veteran Project. *Military Medicine, 150,* 571–576.

Pelto, P. J. (1970). *Anthropological research: The structure of inquiry.* New York: Harper & Row.

Pike, K. (1954). *Language in relation to a unified theory of the structure of human behavior* (Vol. 1). Glendale, CA: Summer Institute of Linguistics.

Podolefsky, A., & Brown, P. J. (1999). *Applying anthropology* (5th ed.). Mountain View, CA: Mayfield Publishing.

Powell, R. A., Single, H. M., & Lloyd, K. R. (1996). Focus groups in mental health research: Enhancing the validity of user and provider questionnaires. *International Journal of Social Psychiatry, 42,* 193–206.

Reckase, M. D. (1996). Test construction in the 1990s: Recent approaches every psychologist should know. *Psychological Assessment, 8,* 354–359.

Resnick, H. S., Kilpatrick, D. G., & Lipovsky, J. A. (1991). Assessment of rape-related posttraumatic stress disorder: Stressor and symptom dimensions. *Psychological Assessment, 3,* 561–572.

Schwab, D. P. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior, 2,* 3–43.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23,* 153–158.

Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7,* 300–308.

Soclari. (2003a). The Ethnograph (Version 5.0 for Windows) [Computer software]. London: Sage.

Soclari. (2003b). *QSR N6.* [Computer software]. London: Sage.

Solomon, Z., Mikulincer, M., & Hobfoll, S. E. (1987). Objective versus subjective measurement of stress and social support: Combat-related reactions. *Journal of Consulting and Clinical Psychology, 55,* 577–583.

Spencer, L., Faulkner, A., & Keegan, J. (1988). *Talking about sex.* London: Social and Community Planning Outreach.

Stretch, R. H., Vail, J. D., & Maloney, J. P. (1985). Incidence and etiology of posttraumatic stress disorder among Army Nurse Corps Vietnam veterans. *Journal of Consulting and Clinical Psychology, 53,* 935–936.

Suen, H. K. (1990). *Principles of test theories.* Hillsdale, NJ: Erlbaum.

Switzer, G. E., Wisniewski, S. R., Belle, S. H., Dew, M. A., & Schultz, R. (1999). Selecting, developing, and evaluating research instruments. *Social Psychiatry Psychiatric Epidemiology, 34,* 399–409.

Tilden, V. P., Nelson, C. A., & May, B. A. (1990). Use of qualitative methods to enhance content validity. *Nursing Research, 39,* 172–175.

Van Gerwen, L. J., Spinhoven, P., Van Dyck, R., & Diekstra, R. F. W. (1999). Construction and psychometric characteristics of two self-report questionnaires for the assessment of fear of flying. *Psychological Assessment, 11,* 146–158.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost–benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15,* 478–495.

Appendix A

Final Definitions of Constructs

Construct	Definition
Combat	Exposure to stereotypical warfare experiences such as firing a weapon, being fired on (by enemy or friendly fire), witnessing injury and death, and going on special missions and patrols that involve such experiences. This war-zone factor refers to objective events and circumstances and does not include personal interpretations or subjective judgments of the events or circumstances.
Aftermath of battle	Exposure to the consequences of combat including observing or handling the remains of civilians, enemy soldiers, U.S. and allied personnel, or animals, dealing with POWs, and observing other consequences of combat such as devastated communities and homeless refugees. This factor is also conceptualized as cataloging more objective war-zone events and circumstances.
Perceived threat	Fear for one's safety and well-being in the war zone, especially as a response to potential exposure to circumstances of combat including NBCs (nuclear, biological, or chemical agents), missiles, and friendly fire incidents. This factor reflects emotional or cognitive appraisals of situations that may or may not accurately represent objective or factual reality.
Difficult living and working environment	Exposure to events or circumstances representing repeated or day-to-day irritations and pressures related to life in the war zone. These personal discomforts or deprivations may include the lack of desirable food, lack of privacy, poor living arrangements, uncomfortable climate, cultural difficulties, inadequate equipment, and long workdays.
Concerns about life and family disruptions	Worries that deployment might negatively affect other important life domains. These include both career-related concerns (e.g., losing a job or missing out on a promotion, perhaps especially important for members of the National Guard and Reserves) and family-related concerns (e.g., damaging relationships with spouse or children or missing significant events such as birthdays, weddings, and deaths).
Sexual harassment	Exposure to unwanted sexual touching or verbal conduct of a sexual nature from other unit members, commanding officers, or civilians in the war zone that creates a hostile working environment.
General harassment	Exposure to harassment that is nonsexual but that may occur on the basis of one's biological sex or minority or other social status and that is used to enforce traditional roles, or in response to the violation of these roles. Categories of harassment include indirect resistance to authority, deliberate sabotage, indirect threats, constant scrutiny, and gossip and rumors directed toward individuals.

Appendix B

Focus Group Guide

Introduction/General Question

1. Could each of you please introduce yourself and *briefly* describe your assignment and duties during the Gulf War?

War-Zone Preparedness

2. We are interested in your unique Gulf War experience. Please think back to the time just before you were deployed to the Gulf region and were getting ready to go to war.

a. How well did the training you received prepare you for what you encountered at war (for example, the training you received about how to use equipment)?

b. How well do you think that you were prepared for what it would be like when you got to the Gulf region? Prompt: How did your expectations about the war compare with your actual experiences?

Deployment-Related Experiences

3. Now think about what your life was like in the Gulf War region. What is the first thing that comes to mind about your experience in the Gulf War? Prompt: Please describe some of the positive and negative experiences of the war.

Ask the following questions only if they have not been addressed by previous questions:

Combat Exposure

I have heard that the combat experiences of Gulf War veterans were very different from previous wars. Can you please briefly describe some of your combat experiences?

Aftermath of Battle

One of the obvious consequences of war is the aftermath of battle. For example, some soldiers had to deal with the remains of civilians, soldiers, and animals. Others had to deal with POWs. Did you have any of these experiences? What was this like for you?

Perceived Threat

Even though there didn't seem to be as much direct combat in the Gulf War, we know that there was the constant threat of warfare. What was that like for you? Prompt: Some people have said that the anticipation of combat can be as difficult as combat itself. Was this true in your experience?

Difficult Living and Working Environment

Some soldiers have reported that they were exposed to extremely poor living and working condition during the Gulf War. Were you? What about being in the Gulf region was particularly bothersome? What was this like

for you? Prompt (if they mention negative impact): What other kinds of things did you find stressful about the day-to-day life at war?

Deployment Social Support

I know that you had to work very closely with others in the Gulf War. What was that like for you? What were your relationships like with other soldiers in your unit? With your commanding officers? How important were these relationships to you?

Sexual Harassment

One of the unique things about the Gulf War was that many more of the soldiers were women than ever before. Did you work with women and what was that like for you? Prompt: In the past couple of years there has been a lot of media attention on sexual harassment in the military. Do you think that this was a problem in the Gulf War?

Concerns About Life and Family Disruptions

What was it like to be away from your family and life back home while you were in the Gulf War? How did being away from home affect you? How did you communicate with friends and family back home while you were in the Gulf War region? Was it easy or difficult for you to stay in touch? Prompt (if they mention worries about things back home): What things about life back home did you worry about? How did this affect you?

Reentry/Postdeployment

4. Now think back to the time after you came back from the war. What was it like for you to adapt to being at home? Were there things that made this process easier or harder? What were these things? How long did it take for things to get back to normal at home?

Health/Exposure to Chemical or Biological Agents

5. What has your physical and emotional health been like since returning from the war? Prompt (if they mention health problems): Please describe these health problems. Do you think any of these health problems are related to your service in the war? If so, how?

Wrap-Up/Closure

6. Overall, how would you say that your experience in the Gulf War has impacted you? What would you say are some of the positive and negative consequences of your experience in the Gulf War?

Note. Only a subset of these war-related factors is described in this article.