

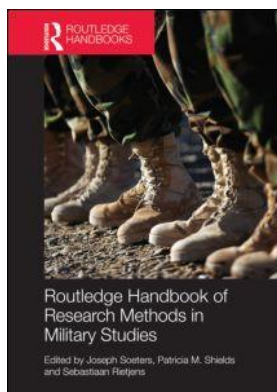
This article was downloaded by: 10.3.97.143

On: 02 Oct 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



Routledge Handbook of Research Methods in Military Studies

Joseph Soeters, Patricia M. Shields, Sebastiaan Rietjens

Survey Research in Military Settings

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9780203093801.ch16>

James Griffith

Published online on: 09 Jun 2014

How to cite :- James Griffith. 09 Jun 2014, *Survey Research in Military Settings from: Routledge Handbook of Research Methods in Military Studies* Routledge

Accessed on: 02 Oct 2023

<https://www.routledgehandbooks.com/doi/10.4324/9780203093801.ch16>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

PART III

Quantitative methods

This page intentionally left blank

16

SURVEY RESEARCH IN MILITARY SETTINGS

James Griffith

J. Griffith (1995) 'The Army Reserve soldier in Operation Desert Storm: Perceptions of being prepared for mobilization, deployment, and combat,' *Armed Forces & Society* 21: 195–215.

In the mid-1970s, the Total Force policy shifted the active components' combat support and service support to the reserve components. This policy made it necessary to mobilize and to deploy reserve component forces in the event of a large-scale war. During the Persian Gulf War, 228,000 reservists were called up, of which 139,207 were Army Reservists. The war, then, provided a test of the Total Force policy: Were the Army reserve components ready for immediate mobilization and deployment in support of combat missions?

The study obtained survey responses from panels of deployed ($N = 259$) and nondeployed ($N = 576$) Army Reserve junior-ranking enlisted soldiers, before and after Operation Desert Storm (ODS). Soldiers were part of a stratified-random, proportional sample of Army Reservists, constructed to adequately represent responses of all deployed and nondeployed soldiers before and after ODS. Survey data were subsequently weighted to represent the responses of all reservists.

Perceptions of equipment and unit leadership preparation showed significant moderate and positive relationships to perceptions that the soldier and the unit were prepared to fight. Favorable spouse and employer attitudes toward reserve service (as reported by the soldier) were significantly and positively related to the soldier's stated intent to stay in the Army Reserve, and negatively related to problems caused by extended mobilization and deployment periods. Soldiers' perceptions of how well their unit leadership was prepared and how well weekend drill prepared them for war were significantly and positively related to reporting for mobilization and staying in the Army Reserve. In contrast, both before and after Operation Desert Storm, large percentages of deployed and nondeployed soldiers reported problems in unit leadership, preparation in individual job and common soldier skills, and weekend drill personnel utilization.

Surveys in the military: Then and now

One of the first dedicated survey efforts was that of Sam Stouffer and his colleagues during World War II (Stouffer et al. 1949) – often considered a classic study of military life supported by data largely obtained through systematic surveys of soldiers. Stouffer and other, now familiar names in psychology surveyed over a half million American soldiers on topics, such as racial integration, officer leadership, unit morale, perceived individual and unit readiness, and others. Survey findings were the basis of several personnel policies, including decisions regarding which units were best suited to land during the D-Day invasion (Converse 2009). Since then, surveys have evolved and become commonplace both in the public and in the military sectors (Kraut 1996). Results of surveys serve to inform leaders and policymakers for a variety of purposes, such as obtaining accurate information regarding need and preferences, evaluating use and effectiveness of programs, and determining what and how to improve organizations (Edwards et al. 1997). This chapter serves as a broad overview of the survey process applied to policy issues of concern to the military. The content of this chapter is structured to respond to key questions, which correspond to essential steps of the survey process.

What is a survey?

A survey may be defined as a *standard method of collecting* information on individuals through the *questioning* of identified *samples of individuals* (Rossi et al. 1983). This definition points to several important elements of a survey – the why, the what, the who, the how and the when, which are used to construct this chapter. “The why” asks, What is the purpose of the survey and potential uses? What is the problem being investigated? “The what” asks, Given the purpose of the survey, what are the relevant domains to represent in the survey and to develop specific survey items? To what extent does the research literature help elaborate on the issue of interest, suggesting specific content and/or a cause map to develop content and to suggest an analysis design? “The who” asks, What is the target population to be surveyed? Who should provide responses to help elaborate on the problem for which the survey is designed? How will potential respondents be chosen? What sampling methods will be considered? “The how and when” asks, What method of data collection and respondent follow-up will be used? What method is most appropriate given the topic, sampling method, and population considered – paper-and-pencil survey questionnaires, telephone survey, web-based survey, and in-person survey? After having conducted the survey, pertinent questions are: What is to be done with the survey responses? How will the results be structured for presentation to sponsors of the survey?

The why, which defines the what: Content domains for the survey

Key to determining what will be gathered in the survey is describing the purpose of the survey. First, the generic purpose of the survey might be described – whether for purposes of providing a point estimate (e.g. what are soldiers’ attitudes toward allowing homosexuals to serve openly in the U.S. military?), monitoring estimates over time for trends (e.g. what are mental health problems of deployed soldiers from year to year?), or evaluating organizational policies and programs (e.g. to what extent do suicide prevention programs impart knowledge and attitude change among soldiers?). Second, specifying the purpose of the survey in very concrete terms is useful. Let us illustrate this beginning step of survey purpose with the illustrative study.

The primary purpose of this survey was to examine U.S. Army Reserve soldiers' readiness to be mobilized and to be deployed. Readiness had to be defined in terms of responses to survey questions. Previously published studies (Reference Note 5 in the illustrative study – Gal 1986; Gal and Manning 1987; Griffith 1988; Hauser 1980; Kellett 1982; Segal and Harris 1993) suggested three broad human dimensions relating to soldier perceptions of combat readiness: (1) the quality of unit leadership; (2) the condition of and soldier familiarity with individual weapons and major weapon systems; and (3) non-problematic family and domestic life. These broad content domains then had to be described in greater detail in order to develop specific survey items and questions.

Methods for identifying what to ask

There are *several methods for developing explicit survey content* for these domains (Edwards et al. 1997), specifically: past surveys, published literature, official documents, interviews of key informants, and focus groups. For the illustrative study, *past Army surveys* were examined for defining and then developing survey items relating to mobilization readiness and then organized by general content domains, such as, leadership, individual and unit training, equipment availability, functionality, and familiarity, etc. Automated *literature searches* (e.g. EBSCO) resulted in more defined content to represent on survey of the illustrative study, including: weekend drill mobilization preparation (e.g. effective use of drill time, spend little time on unnecessary things, train as a team, train soldiers in individual job and common soldier skills); unit leadership mobilization preparation (e.g. trust and confidence in unit officers/NCOs, leaders treat me as a person, apply discipline fairly, care about soldiers, provide good supervision, train as a team, promote teamwork and cooperation), etc. *Official documents*, such as memoranda, policy letters, program descriptions, etc., can be used to gain background on the problem for which the survey will gather data. In the illustrative study, the survey team examined memos and policy letters that described the nature and extent of the problem and probable causes, incentive programs to get youth to join and to remain in reserve military service, etc. Developed survey content included reasons for enlistment and reasons for re-enlistment and the role of the various incentive programs in their decision processes (e.g. knowledge and use of incentives).

Another source for identifying survey content is *interviewing of key informants* on the topic being investigated. These interviews are a good source for broad specification of the topic and related content, or “the cause map” – a pictorial or conceptual understanding of antecedents, intervening variables, and consequences pertinent to the key content domain for which survey data is wanted. To prepare for these interviews, the survey team should prepare a statement regarding the purpose of the survey and broad questions relating to the survey topic, for example, How prepared are reservists to be mobilized and deployed to combat areas? Subsequent questions should cascade to greater specificity, for example, what are areas reservists are most and least prepared? During the interview, the survey team should use probes to embellish areas and seek potential antecedents, for example: “Tell me more about this . . .” and “Why do you think this is the case . . .?” In the illustrative study, interviews were conducted with key personnel, such as the retention and readiness staff at the Continental U.S. Army Commands and several major U.S. Army Reserve Commands. Another form of interviewing is the *focus group*. In this method, participants – usually comprised of 6 to 12 people of the population to be surveyed – are asked questions for purposes of group discussion (O’Brien 1993). First, a few broad questions are asked followed by probes and participants’ elaboration. Responses lead to identifying major themes and subthemes, in addition to possible specific survey questions, items, and response options. For the illustrative study, several focus groups were conducted at several major Army headquarters. Two survey team members facilitated the discussion of 10–15

junior-ranking enlisted (privates through sergeants) for one and a half hours concerning reserve military service: reasons for joining and leaving; readiness for mobilization and deployment; quality of unit training; availability, quality and familiarity with weapons; quality of unit leadership; and the role of civilian employment and family life in reserve military service. The survey team took notes, identifying content for each domain, and later, developed as survey items organized by identified content domains.

Generating specific items

Specific survey content should be developed by domains identified in the review of previous questionnaires, literature, focus groups, and other initial activities. At most, there should be a dozen or so content domains for the survey. Depending on the specificity desired, content should be developed for each domain. Theories, models, and organizing frameworks identified are useful in developing explicit survey content, as well as suggesting an analysis plan. Ajzen and Fishbein's (1980) theory of reasoned behavior, for example, served as an organizing perspective for several large-scale military surveys, including the Army Communications Objectives Measurement Study (Rhoads and Elig 1988), which assessed the effectiveness of various marketing and advertising strategies to recruit soldiers for the U.S. Army's all-volunteer force. Survey content inquiring about information from respondents take two general forms: *questions* and *statements*. Examples of each are found in Table 16.1

Questions ask for either one or multiple responses, respectively: What was your age at your last birthday? Or, what are the reasons you joined reserve military service? Response options can take two general forms: *open-ended*, which allow the respondent to write in text; and *closed-ended*, which provide predefined response options for the respondent. While the former allows more spontaneity and at times more comprehensiveness, such responses also entail coding – that is, developing a scheme to organize responses into general categories and assign numbers to the responses for analysis. The latter, closed-ended response options, while limiting respondents' answers, provides easy data recording of responses and is preferred, in particular, for surveys of small budgets. Closed-ended responses can take many forms (as displayed in the table). Of particular note is whether response options indicate from positive to negative versus much quantity to none (see Table 16.2). These descriptors can then be arranged as a Likert-type response scale. Response options include both descriptors and the corresponding numerical assignments (ordinally scaled, e.g. from low to high or high to low). Likert-type scales have anywhere between 3 to 7 responses options, though, generally, the variance sufficient for analyses provided by few point response options (usually minimally, three) is similar to that provided by more response options (Dawes 2007; Garland 1991). Other scales for attitudinal assessment are discussed in Oskamp and Schultz (2005), namely, Bogardus Social Distance Scale, Thurstone's Equal-appearing Interval Scale, Guttman's Scalogram.

These response scales are then used to respond to specific questions or statements contained in the survey instrument. Simple "one-line" questions or statements for respondents are preferred, as it takes less effort for respondents to read and answer. Insofar as possible, the survey team should develop questions and statements of similar format, for example, a series of statements that require the same response options. This format, again, makes it easy for the respondent to read and respond. Table 16.3 outlines common problems in crafting survey items.

Some questions or statements may require more than one response, such as reasons for joining reserve military service. Responses to such items may represent either independent responses, i.e. coding each response as yes versus no, or dependent responses requiring representation of various combinations. The first situation is considered mutually exclusive responses, either yes

Table 16.1 Examples of survey questions and statements

Questions

What is your military status? MARK ONE:

- M-day soldier, only “part-time”
- Full-time soldier on temporary funding
- Technician, e.g. ADSW
- Full-time AGR soldier
- Other

In the last year, have you been deployed to any one of the following locations?

MARK ALL THAT APPLY:

- Iraq
- Afghanistan
- Kuwait
- Other location

Statements

For each statement, place a checkmark to the right of each statement indicating your response.

<i>Survey statement</i>	<i>Strongly agree</i>	<i>Agree</i>	<i>Disagree</i>	<i>Strongly disagree</i>
There is a lot of teamwork and cooperation among members in my unit				
When things don't get done, members of my unit pull together				
Soldiers in my unit stick together to accomplish the mission				

Table 16.2 Common descriptors for response options

<i>Types of response</i>	<i>Example descriptors assigned to numerical ratings on Likert-type scale</i>	<i>Ratings represent</i>
Agreement	Strong agree, agree, disagree, strongly disagree	Positive versus negative
Satisfaction	Very satisfied, satisfied, dissatisfied, very dissatisfied	Balanced
Quality	Very good, good, average, poor, very poor	Positive versus negative
Expectations	Much better than expected, better than expected, as expected, worse than expected, much worse than expected	Positive versus negative
Effectiveness	Very effective, effective, ineffective, very ineffective	Positive versus negative
Likelihood	Very likely, likely, unlikely, very unlikely	Positive versus negative
Frequency	Always, often, seldom/rarely, never	More to less
Extent	To great extent, to some extent, to small extent, to no extent	More to less
Importance	Very important, somewhat important, slightly important, not all important	More to less

Table 16.3 Common problems in crafting survey items and helpful hints

<i>Common problem</i>	<i>Example of problem</i>	<i>Helpful hint</i>
Use of double-barreled items	I am in the military to serve my country and feel an obligation to others	Be sure to keep survey item to one referent
Use of complex language, not appropriate for population being surveyed	My military services are a deterrent to the global spread of totalitarian regimes	Keep language simple and appropriate for population being surveyed
Being vague regarding referent behavior and related factors	Are you satisfied with military service?	Be specific regarding referent behavior and related factors
Use of leading questions	Most people feel that serving in the military is. Do you agree?	Avoid questions which introduce bias in response
Use of double negatives	I do not like the idea of not receiving bonuses	Do not use two negative in survey item

or no, to each item. The second situation is when responses are not mutually exclusive and can take various combinations, usually making coding and analysis more difficult. That is, there can be any number of various combinations of responses that will have to be coded as unique responses. Additionally, response options that often pose difficulties for the respondent and survey team are: Don't Know, Not applicable, Neutral, and Other. Don't knows and/or Not applicable are often seen as reflecting a neutral position in an attitude assessment, as responding "neutral." Don't know and Not applicable responses may, however, reflect a non-attitude, or not even responding to the response scale or item. For these ambiguities, survey research studies suggest omitting a middle response category for attitude assessment.

Determining reliability and validity

Survey items nearly always serve to measure a variable or construct, and thus, respondents' answers to such items must show some degree of reliability and validity. A common method of reliability assessment is examining the consistency in responses to survey items which are thought to assess the corresponding construct. The SPSS Reliability routine can determine internal response consistency – either by way of Cronbach alpha or item–total correlations or both (see Pedhazur and Schmelkin 1991). Factor analysis (Gorsuch 1983) can also determine the extent to which a set of survey items assess the same underlying construct. In confirmatory factor analysis, the researcher has substantial justification for the arrangement of survey items in relation to facets of the underlying construct. The researcher specifies which of the survey items correlate with which facets of the underlying construct. For example, commitment is often thought of as having facets of normative, affective, and continuance (Meyer and Allen 1991). Survey items presumed to assess these facets would be tested simultaneously using confirmatory factor analysis (see SPSS add-on module called AMOS; Arbuckle 2009). If the researcher is less certain about the way in which items arrange themselves in relation to underlying factors, then alternatively exploratory factor analysis can be used (see SPSS Data Reduction, Factor Analysis). Sometimes, it is appropriate to conduct cognitive labs on a limited number of respondents to discern whether the content of questions is understood as intended. An example method is "read aloud," where the investigator reads the questionnaire item to the respondent (Jobe and Mingay 1990). The respondent then repeats back his (her) understanding of what has been said. The investigator can then tell the extent to which the meaning of an item has been conveyed.

A more complicated measurement issue is one of validity, whether survey items obtain data on what they purport to assess. Face and content validity are, perhaps, the easiest to demonstrate. Survey items have face validity if they look like they measure what they are supposed to measure. For example, survey items that ask about various experiences of immediate leaders would seem to assess leadership. Content validity refers to the extent to which a measure represents all facets of a given construct. Thus, measures of leadership would have to include various aspects identified in the literature – from theories and/or empirical findings. Several other methods to establish validity of measures include concurrent, predictive, and discriminant. For an introduction to these topics, see Oskamp and Schultz (2005).

Arrangement of survey items

After survey questions and statements have been drafted along with respective response options, the content needs to be arranged in a survey questionnaire. Items of similar content should be grouped together, again, for ease of the respondent's understanding and responding. More simple, less threatening content should be placed at the beginning of the questionnaire and more complex, obtrusive content to the end of the questionnaire. In this way, the respondent's trust and confidence is built as they go through the questionnaire, and thus, are more likely to respond to more personal questions later, such as age, income, etc. The survey team should be mindful of common problems encountered in respondent's responses. Respondents often answer survey content without regard to content, for example, check all items as "agree." Presenting survey content as alternating positive and negative content prevents this problem (called a "balanced format"). Respondents also may respond according to what they believe the survey team wants then to say, called "response acquiescence." Finally, respondents may wish to present themselves in the best light and answer accordingly; this is called social desirability. Each of these problems results in responses that are not genuine, and thus, threatens the validity of the measurement and should be avoided. It is advisable to examine the responses of the first 100 or so questionnaires to determine whether such problems exist for remedial action.

The who

"The who" asks the questions, What is the target population to be surveyed? Stated alternatively, who should provide responses to help clarify the problem for which the survey is designed? And subsequently, how will potential respondents be chosen? What sampling methods will be considered? Ideally, everyone in the target population would be surveyed. But, due level of effort constraints – both staff and money, this is rarely feasible. This would entail providing a survey instrument to everyone with subsequent follow-up to ensure every respondent responds. It is more realistic to survey a smaller subset of the target audience with repeated follow-up to ensure high completion rates. And indeed, regardless of the population's size, a carefully chosen sample of a subset numbering 1,500 achieves a reasonable degree of precision for estimates (i.e. the population percentage is expected to plus or minus 3 percent of the sample estimate). Sampling requires knowing eligible individuals in the entire population followed by systematic selection of respondents for the sample (Fowler 2009). Before beginning, several terms are important to understand when sampling: universe or population, sampling frame, sample, precision, and error.

The *universe* or *population* is a term used when referring to all possible respondents. For example, a survey of U.S. Army soldiers would be all soldiers currently serving in the U.S. Army. The *sampling frame* refers to a listing of all possible respondents and is useful in selecting a smaller subset for purposes of surveying them. Such lists often do not include every eligible individual

in the population, though nearly everyone. For example, in the earlier example study, not every soldier serving in the U.S. Army may be found on a personnel list, due to being newly accessed.

The smaller subset from the sampling frame is called the *sample*. The number of members chosen from the population determines the precision and error associated with any estimate derived from the survey data. Small sample sizes (up to about 500) generate estimates having more error and less precision. For example, the “true” population percent associated with a sample percent (derived from a sample of 500) would be between plus or minus 5 percent of the sample percent. Error and precision are important when wanting to detect differences between groups (men versus women, junior-ranking enlisted versus others, those treated or not). A convenient way to calculate approximate error associated with a given sample percent is:

$$\sqrt{(p \star q / N)} \star 1.96 \star 100 = \text{+/- band within which the population parameter is likely to fall in 95 out of 100 random samples.}$$

Let $p = 0.5$ (where maximum variability occurs for a proportion); $q = 1 - p$; and $N = \text{sample size}$.

1.96 is the z-value associated with 95% confidence interval, and the multiplier of 100 converts the proportions into percentages.

In determining the sample size, the survey research team should ensure enough precision is achieved to show differences between groups (demographically or programmatically defined) exceed expected error. Power analysis is used to calculate sufficient sample sizes to allow for this. (For more detail, see chapter 6, “Determining Sample Size,” in Rea and Parker 1992.) Sampling design also reflects policy interests. For example, there may be interest in knowing certain characteristics of low-frequency groups in the population. Thus, such groups would have to be over-sampled or over-represented in the sample in order to yield a reasonably reliable estimate on survey responses, and when combined with all respondents would have to be weighted “downward” to represent their occurrence in the population.

There are several ways to determine who is included. A major concern in survey research is implementing a design that obtains results from a smaller subset of the population (called a sample), which represents, to some degree of confidence, population results sample. Broadly, sampling occurs as two approaches: *probabilistic* and *non-probabilistic*. The non-probabilistic approach is problematic in that it is not possible to determine whether the sampled respondents actually represent the larger population. It is more likely the probabilistic approach will achieve this, that is, a representative sample that resembles the population in key background characteristics, especially those characteristic correlated with the topic of interest. This is *external validity* or the extent to which results obtained from a sample describe those obtained from the entire population.

To achieve such results, the sample must be a probabilistic sample (simple random, stratified-random, and systematic, etc.) Key statistical methods also rely on probabilistic sampling, such as error band associated with statistics derived from the sample of respondents and inferential statistics used to determine whether associations among survey variables occur by chance (i.e. sampling and measurement errors) or represent true observed associations. Probabilistic surveys require some way to enumerate all potential members of the population. That is, lists of potential respondents are available and/or members are nested in some way to sample from the “nestings” to achieve randomness, such as households or telephones. The *non-probabilistic* approach to sampling is also called opportunistic sampling. While some criteria for inclusion in the sample may have been applied, not everyone has been enumerated and assigned a chance

of being selected as a respondent. Both probabilistic and non-probabilistic approaches occur in several forms, as described below.

Probabilistic sampling methods

To collect a *simple random sample*, each individual of the target population is assigned a number. A set of random numbers is then generated and the individuals having those numbers are included in the sample. To perform *systematic sampling*, individuals in the target population must first be arranged in some ordering scheme. Individuals are then selected at regular intervals through that ordered list. A random start on the list is determined and then every n th individual is chosen for the sample. This method assumes that there are no biases in the arrangement of cases on the list (called periodicity). For example, choosing every n th house along a street might result in taking too many corner households, which typically have bigger lots and more expensive houses. In *stratified-random sampling*, respondent characteristics important to the survey estimates are considered, such as rank, gender, etc. Population members are then arranged into cells corresponding to these characteristics, called strata. Members are sampled randomly from the various cells (private–junior sergeant men, privates to sergeant women, senior sergeant men, senior sergeant women, etc.) Members can be sampled proportional to their occurrence in the population (proportional probability sampling) or not (disproportional probability sampling). Disproportional probability sampling would be appropriate if reliable estimates are needed for a low-frequency group, such as senior sergeant who are women. This then would provide more reliable estimates from the survey (estimate with smaller error band). If, however, overall estimates for the population are desired, strata from which members were disproportionally sampled would have weights applied. Weights are derived as a ratio of rate of occurrence in population/rate of occurrence in the sample. The weights are used to multiply the cases having the sampling characteristics, e.g. senior sergeants who are women.

Cluster sampling is used when it is either impossible or impractical to compile an exhaustive list of individuals who comprise the target population. Sometimes lists of members of a population are not readily available. But, where the members typically occur or “reside” is known. That is, population members are known to “cluster” or to gather in geographically defined groups, such as by their unit membership (e.g. people reside in households, census tracts, cities, etc.). With cluster sampling, it is best when the clusters are internally heterogeneous on the characteristics being studied, so as not to bias results on any particular respondent characteristic. Also, smaller numbers in clusters are desired. The larger the number of respondents per cluster, the greater the within-group variance making the sampling method less efficient.

Because military personnel situate themselves as intact groups, obtaining a representative sample is generally accomplished through cluster sampling where soldiers within each randomly selected unit or cluster would be surveyed. A recent example of cluster sampling was employed by Schaubroeck et al. (2012) where brigades were chosen from combat divisions, battalions from brigades, companies from battalions, platoons from companies, and finally, squads from platoons. Soldiers in sampled squads then completed surveys. Another recent examples are the Army’s Mental Health Advisory Team Surveys (MHAT 2008) where units (squads and platoons) were randomly selected from larger organizational clusters, such as brigades and battalions. Clustering can reduce travel and administrative costs of the survey by not having to travel to all units deployed throughout Afghanistan and surveying only a few in each unit (as in simple random sampling of soldiers).

Non-probabilistic sampling methods

Non-probabilistic sampling occurs in several types; convenient, purposive, and quota. A *convenience sampling* is a matter of taking whoever is eligible to complete the survey. Volunteers would constitute a convenience sample. In *quota sampling*, respondents are taken from the population in some pre-specified numbers so as to achieve some heterogeneity in the sample. However, the selection of respondents is not random. A *purposive sample* is having chosen a non-representative subset of a larger population for a very specific need or purpose.

Sampling in the illustrative study

The illustrative study (presented earlier) used a fairly complicated sampling design, which required several professional statisticians to develop, review, and apply final weights to the sample. The sample was a stratified-random, disproportional sampling. The study reported survey data gathered in 1991 and in 1992. Previously, the survey had been conducted annually starting in 1988. A list of reserve soldiers was obtained from the Army's personnel system (called the Standard Installation Division Personnel Reporting System). Each year, the sample numbered about 31,000 reservists who were randomly selected from strata defined by background characteristics, such as rank, gender, and race. Selected respondents represented were members of about 3,300 reserve units across the U.S. In the first three years of the survey, annual samples were cross-sectional. In the fourth year of the survey, the First Gulf War broke out. Senior Army leaders were interested in the level of preparedness of reservists for the war, in addition to their experiences and how these related to their intentions to remain in reserve military service. Given this policy interest, the annual cross-sectional sample was augmented by reservists who had been deployed or not, to allow for comparisons between the two groups concerning readiness, deployment experiences, and retention intentions. This sampling also provided panels of reservists from the previous year to the current year, which had been deployed or not. The panels provided survey responses on the same soldiers before and after mobilization and deployment. To accomplish the augmented sampling, reservists were stratified by whether they had been deployed or not. Deployed reservists were over-sampled to ensure adequate numbers for point estimates and analyses. Both deployed and nondeployed reservists were stratified into several groups: nonmedical personnel and medical personnel (including physicians, licensed nurses, and practical nurses). Many medical personnel deployed in support of the First Gulf War. There was thus interest in examining medical personnel's intention to stay or leave reserve military service upon return. Both nonmedical and medical personnel were then stratified by rank, gender, and minority status due to the policy interests in junior-ranking, women, and minority soldiers.

The how and when of survey data collection

"The how and the when" asks, What method of data collection and respondent follow-up will be used? What method is most appropriate given the topic, sampling method, and population considered? There are three primary modalities for gathering survey data: paper-and-pencil surveys (either group or individually administered), through personal interaction (individually in personal interview or over the telephone), and via the Internet (Groves et al. 2009).

Paper-and-pencil surveys take the form of hard-copy material on which the respondent records their answers, administered either in group settings or mailed individually to prospective respondents. Group administration requires survey staff to travel to the site of administration. This practice can achieve a high completion rate due to the presence of the survey team, unit

leadership, and others in the unit, in addition to command allocation of specific time. The flexible individually administered surveys allow the respondent to complete the survey but usually on their own time. Unfortunately these surveys are less likely to be completed. Mail surveys often have long delay times between delivery and eventual return of the completed questionnaires. Individually administered surveys often require substantial follow-up to get respondents to complete the survey questionnaires. Follow-ups should usually occur one week to ten days after initial contact. Two to three follow-ups may be required to achieve an acceptable completion rate (usually strive to obtain 70 percent or higher). While the absence of interviewers during data collection avoids interviewer bias, there is little opportunity to clarify ambiguous survey questions and items.

Telephone interviewing is another modality of gathering survey responses (Dillman 2006). Telephone surveys more often occur as computer-assisted telephone interviewing (called CATI), where interviewers read screens from a pre-programmed survey instrument and enter respondent's answers directly into the computer program. Such telephone surveys are usually fairly expensive involving a large staff who call from telephone call centers. A less expensive method is the interactive voice response system, which involves pre-recording of the survey questions and responses, and the respondent answers questions by saying the response or touching a key pad on the telephone. There are several advantages to telephone surveys. Interviewers can actively encourage sampled individuals to participate in the survey, leading to higher response rates. By answering questions respondents have about the meaning of questions and response options, interviewers can clarify and increase comprehension. Disadvantages of telephone surveys include: time and expense in training interviewers, development of the computer-assisted survey screens, and potential for interviewer influence on respondents' answers.

A modality growing in popularity is the *web-based survey* (Bethlehem and Biffignandi 2012; Vehovar and Lozar 2008). Such surveys are initiated by informing prospective respondent (via email or mail) of a web site where he (she) is to go to complete a survey. Web surveys are relatively inexpensive once set-up. There are several advantages of web-based surveys. The data collection period is shortened because respondents are typically notified electronically (via email) and their responses are recorded instantaneously as they respond. Web-based data collection is less intrusive, allowing sensitive questions to be asked and minimizing interviewer influences on respondents' answers. Web-based surveys also allow for – complex skip patterns, implemented unbeknownst to the respondent, pop-up instructions to help clarify questions and response options, and lists of answer choices as drop downs. Distinct disadvantage of web-based surveys are the availability of Internet-connected computers as well as respondents' abilities to use computers and navigate the Internet.

In-person interviews are the final form of data collection described. Here, interviewers are recruited and trained. Interviewers also undergo extensive training in the survey content – questions and response options, and how to record responses uniformly across respondents. Interviewers then go to where the prospective respondent is physically located to conduct the interview. Nowadays, interviewers generally use personal computers, which display questions on successive screens (called computer-assisted personal interview or CAPI). The interviewer reads them to the respondent, and then enters the respondent's answers, questions and response options. The interviewer record respondents' answers which are then uploaded regularly uploaded to mainframe computers housing all respondents' answers. A major advantage of the personal interview is personal contact with respondents, which results in high cooperation and the lowest refusal rates. The interview also allows for longer, more complex questions and responses. Disadvantages include the high costs of interviewers personally contacting selected respondents, the longer data collection period, and the potential for interview influence on respondents' answers.

Whatever the modality, the importance of achieving a high response cannot be over-emphasized. While mail surveys and Internet surveys are less expensive, usually require much follow-up to achieve acceptable response rate. More person-contact methods – telephone or in-person interviews – achieve higher response rates. The survey team should develop a plan for follow-up whatever the modality of data collection. The planning includes a method to track and trace survey questionnaires to determine whether the respondent responded and for those who have not to follow-up with reminders. Knowing who responded and did not also allows for adjustments to the data set once data collection is closed. For example, having basic demographic characteristics for responders and nonresponders allows for the possibility of weight adjustments, i.e. for those respondents of specific characteristics who under-responded, the respondents who did answer might be weighted upward to represent them proportional to their occurrence in the sample.

Data collection method in the illustrative study

In the illustrative study, the survey sample consisted of about 31,000 reservists scattered across the U.S. Telephone numbers were not readily available, and if they were, the cost associated with telephone survey would be prohibitive, though likely to achieve a high response rate. In-person interviewing was also cost prohibitive – travel costs for survey staff visiting many locations and spending lengthy time periods collecting data. Mailing questionnaires to respondents was less expensive. Questionnaires were designed and printed as optical scannable forms, with identification codes. Each respondent was then matched to a code. In that way, responding and nonresponding reservists could be tracked for follow-up. It further facilitated the weighting of data set to more equitably represent under-responding reservist. Several reminders were sent out during the months of data collection. In follow-ups, nonresponding reservists were sent replacement questionnaire forms. Slightly over 41 percent of the sample completed and returned questionnaires. Another important aspect of data collection, especially before ending data collection, is examining the extent to which groups within the sample responded at similar rates. If not, then there may be problems in non-representativeness of the survey responses. In the case study, completion rates were compared across the sampling strata (gender, rank, and deployments status) and the percentage of personnel in each stratum in the initial sample was compared to the percentage that completed surveys. The percentage of personnel initially sampled who completed surveys by geographic location of the unit, U.S. commands, regional commands, and states, were compared. No systematic difference between the percentages of personnel who had been initially sampled and who had then completed surveys among the various geographic and individual characteristics were observed. If systematic differences are observed between respondents and non-respondents in known characteristics for each group, such as age, rank, gender, race, etc., then weights can derived to “weight up” under-responding groups and “weight down” over-responding groups. One method for weighting is the percentage represented doing population divided by the percentage in the sample. For example, if women represent 20 percent of the population, yet in the probability sample, only 15 percent respond, each responding woman would get a weight of 1.33.

The reporting

After the survey is conducted, pertinent questions are: What is to be done survey to the responses? Specifically, what is the design for the analysis and what results will the design yield? How will the results be structured for presentation to sponsors? How should the purpose and

questions to be answered by the survey results, or how should the research questions, structure the analysis and presentation of findings?

Broadly, results can be presented as simply descriptive, comparative or associative (see chapters 8 and 9, Rea and Parker 1992). Descriptive presentations involve survey results presented for one group, e.g. the responding sample. Comparative presentation involves presentation of survey results for two or more groups. These groups are usually defined by the survey purpose, e.g. comparison of deployed versus nondeployed on readiness survey items. Finally, associative presentation involves relating survey results to important outcomes of interest. For example, regression analyses might be conducted in which perceptions of preparedness, deployment experiences, etc. are regressed on intentions to stay or leave reserve military service. Non-scientific audiences, likely policymakers sponsoring the survey, typically understand presentation of percentages better than presentation of means and other forms of descriptive statistics. In each presentation, it is important to report the appropriate confidence level of the finding or error associated with finding. For descriptive presentations, this is usually an error band or interval of confidence.

The illustrative study used all three approaches to analysis and reporting. Descriptively, percentages of soldiers in the sample who represented various categories in demographic subpopulations were presented. Descriptive analyses also included more complex scaling methods, such as exploratory factor analysis. Items representing various content of readiness (e.g. equipment, unit leadership, training, and family) underwent exploratory factor analysis to examine the extent to which all items could be summarily used to describe dimensions of readiness. The arrangement of items (i.e. items that correlated most with each factor) was used to derive summated, averaged scale scores for equipment, unit leadership, training, and family. Comparatively, responses of deployed soldiers to readiness items were compared before deployment (1990) and after deployment (1991). Similarly, responses of nondeployed soldiers to the same items were compared between 1990 and 1991, noting how these differed to the deployed soldiers' responses over-time. Finally, associative analyses included simple correlations and multiple regression analyses in which associations between the readiness scale scores and several outcomes were examined, such as perceived self-preparation for combat, unit preparation for combat, willingness to report for duty were activated, and intention to remaining remain in reserve military service.

Other important elements of reporting include details on sampling, data sources and analytic approaches. Description of the sampling design should include what the method was and how it was accomplished, and the number of eligible respondents and the number who completed. Summary statistics (percentages) of background characteristics of soldiers in the responding sample are compared to those of the population to provide a sense of respondent sample representativeness. Data source should be described, including the survey questionnaire and how it was developed, pretested, and how variables were constructed, in particular, scale score through factor analysis and reliability of ratings given to items comprising constructed variables or scales; and the use of any ancillary data sets. The analytic approach or design should be described, including the rationale and the steps taken in the analysis – both corresponding to the survey purpose and research questions.

Summary

Surveys are ever-present and expanding in their use in society. So too, the military has increasingly used surveys to gather information from soldiers – their backgrounds, experiences, and attitudes – for purposes of informing the development and implementation of policies. To this end, the present chapter provides an overview of the survey research method. The survey

process is described in response to several key questions – the why (survey purpose), the what (survey questionnaire content), the who (selection of potential respondents), the how and when (data collection method), and finally, how best to summarize and present survey data. To show more concretely how these steps occur, an illustrative study was used throughout.

References

- Ajzen, I. and Fishbein, M. (1980) *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs, NJ: Prentice-Hall.
- Arbuckle, J.L. (2009) *Amos 18 User's Guide*, Chicago, IL: SPSS, Inc.
- Bethlehem, J. and Biffignandi, S. (2012) *Handbook of Web Surveys*, Wiley handbooks in survey methodology, Hoboken, NJ: Wiley.
- Converse, J. (2009) *Survey Research in the United States: Roots and Emergence 1890–1960*, New Brunswick, NJ: Transaction.
- Dawes, J. (2007) “Do Data Characteristics Change According to the Number of Scale Points Used? An experiment using 5-point, 7-point and 10-point scales” *International Journal of Market Research*, 50: 61–77.
- Dillman, D.A. (2006) *Mail and Internet Surveys: The Tailored Design Method*, 2nd edn, Hoboken, NJ: Wiley.
- Edwards, J.E., Thomas, M.D., Rosenfeld, P. and Booth-Kewley, S. (1997) *How to Conduct Organizational Surveys: A Step-by-Step Guide*, Thousand Oaks, CA: Sage.
- Fowler, F.J. (2009) “Sampling,” in *Survey Research Methods*, 4th edn, Thousand Oaks, CA: Sage, 19–47.
- Gal, R. (1986) “Unit Morale: From a theoretical puzzle to an empirical illustration – an Israeli example,” *Journal of Applied Social Psychology*, 16: 549–564.
- Gal, R. and Manning, F.J. (1987) “Morale and Its Components: A cross-national comparison,” *Journal of Applied Social Psychology*, 7: 369–391.
- Garland, R. (1991) “The Mid-Point on a Rating Scale: Is it desirable?” *Marketing Bulletin*, 2: 70.
- Gorsuch, R.L. (1983) *Factor Analysis*, Hillsdale, NJ: Erlbaum.
- Griffith, J. (1988) “The Measurement of Group Cohesion in U.S. Army Units,” *Basic and Applied Social Psychology*, 9: 149–171.
- Griffith, J. (1995) “The Army Reserve Soldier in Operation Desert Storm: Perceptions of being prepared for mobilization, deployment, and combat,” *Armed Forces & Society*, 21: 195–215.
- Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2009) *Survey Methodology*, Hoboken, NJ: Wiley.
- Hauser, W.L. (1980) “The Will to Fight,” in S. Sarkasian (ed.) *Combat Effectiveness: Cohesion, Stress and the Volunteer Military*, Beverly Hills, CA: Sage.
- Jobe, J.B. and Mingay, D.J. (1990) “Cognitive Laboratory Approach to Designing Questionnaires for Surveys of the Elderly,” *Public Health Reports*, 105: 518–524.
- Kraut, A.I. (1996) “An Overview of Organizational Surveys,” in A. Kraut (ed.) *Organizational Surveys: Tools for Assessment and Change*, 1–14, San Francisco, CA: Jossey-Bass.
- Mental Health Advisory Team (MHAT) V (2008) *Operation Iraqi Freedom 06–08*, Office of the Army Surgeon General and U.S. Army Medical Command, Washington, D.C.
- Meyer, J.P. and Allen, N.J. (1991) “A Three-Component Conceptualization of Organizational Commitment: Some methodological considerations,” *Human Resource Management Review*, 1: 61–98.
- O’Brien, K. (1993) “Improving Survey Questionnaires through Focus Groups,” in D.L. Morgan (ed.) *Successful Focus Groups: Advancing the State of the Art*, 105–117, Newbury Park, CA: Sage.
- Oskamp, S. and Schultz, P.W. (2005) *Attitudes and Opinions*, 3rd edn, Hillsdale, NJ: Erlbaum.
- Pedhazur, E.J. and Schmelkin, L.P. (1991) *Measurement, Design, and Analysis: An Integrated Approach*, Hillsdale, NJ: Erlbaum.
- Rea, L.M. and Parker, R.A. (1992) *Designing and Conducting Survey Research: A Comprehensive Guide*, San Francisco, CA: Jossey-Bass.
- Rhoads, M.D. and Elig, T. (1988) *The Army Communications Objectives Measurement System (ACOMS): Survey Methods*, Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Rossi, P.H., Wright, J.D. and Anderson, A.B. (1983) “Sample Surveys: History, current practice, and future prospects,” in P.H. Rossi, J.D. Wright and A.B. Anderson (eds) *Handbook of Survey Research*, 1–20, Orlando, FL: Academic.

- Schaubroeck, J.M., Hannah, S.T., Avolio, B.J., Kozlowski, S.W.J., Lord, R.G., Trevino, L.K., Dimotakis, N. and Peng, A.C. (2012) "Embedding Ethical Leadership within and across Organizational Levels," *Academy of Management Journal*, 55: 1053–1078.
- Segal, M.W. and Harris, J. (1993) "What We Know about Army Families," Technical Report No. DAALO03–86–D–0001, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.
- Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. and Williams, Jr., R.M. (1949) *The American Soldier: Adjustment during Army Life*, Vol. 1 in *Studies in Social Psychology in World War II*, Princeton, NJ: Princeton University Press.
- Vehovar, V. and Lozar, M.K. (2008) "Overview: Online surveys," in N. Fielding, R.M. Lee, and G. Blank (eds) *The SAGE Handbook of Online Research Methods*, 177–194, Thousand Oaks, CA: Sage.