Handbook for Survey Development

U.S. Naval War College War Gaming Department

Douglas R. Ducharme EdD

ABSTRACT

The War Gaming Department at the U.S. Naval War College has been war gaming since 1887. This paper describes how we think about incorporating survey methods into our analysis and survey development steps to consider across each phase of our war gaming process. It augments the War Gamers' Handbook to provide background on analytical war gaming, our terminology, and our research design process.

Citation: Douglas R. Ducharme. War Gaming Department Working Paper WGD_20202, U.S. Naval War College, Newport RI. November 2020.

Dr. Ducharme is a War Gaming Specialist for Network and Simulation Technologies, Inc. (NETSIMCO) in the War Gaming Department supporting analytical war gaming for the U.S. Department of Defense senior leadership.

Douglas R. Ducharme EdD

Table of Contents

Douglas R. Ducharme EdD

Introduction

Purpose of this Handbook

The purpose of analytical war games involves gaining insights into complex problems. In the process of conducting war games, data are collected relevant to game objectives. Development of survey instruments is a key activity in the game management process that provides a means to collect data. The War Gamers' Handbook (Burns et al., 2013) mentions survey use four times, such as for the use of collecting demographic data through a survey, leveraging survey questionnaires to obtain player views on the cell's success or lack of success attaining cell objectives, refining survey questions during the development and testing phases, and that creating survey questions as a project team assignment duty for the game analyst. However, while the War Gamers' Handbook describes what a game analyst does to incorporate survey methods, it fails to explain in detail how to develop, test, and administer survey instruments.

The goal of this handbook is to enable war gamers to prepare and conduct surveys and to help war gamers become better users of survey results. Specifically, this handbook attempts to address how to ask questions, how to collect reliable and valid information, and how to analyze and report results.

Handbook Format

This handbook consists of two parts. Part 1 provides an overview of the fundamentals of survey methods, including the definition and utility of surveys, collection types, data types, and the key steps in conducting surveys. The applicability of survey methods for use in war games will be addressed. Finally, Part 1 includes a thorough discussion of reliability and validity considerations associated with survey methods.

Part 2 provides a step-by-step guide of survey considerations and activities across each game management process phase. It includes how to take the game objectives using a literature review to design a survey plan during the tasking and design phases. The development phase section provides the best practices for drafting survey questions. Tips on testing, rehearsal, and execution of survey methods follow. Finally, a comprehensive section for the analysis phase addresses the various techniques used to assess survey data.

Throughout this handbook, various examples from war games and other survey projects are included to provide practical applications of the concepts described in various sections. The appendix provides additional detailed instructions on various survey method procedures.

Douglas R. Ducharme EdD

Part 1 – Survey Fundamentals

Survey Definition

The term survey refers to the collection of information to describe, compare, or explain the respondents' knowledge, attitudes, and behavior (Fink, 2003a). Survey may often be perceived in narrow terms as a form completed by an individual. However, survey can be perceived in broad terms as any means to collect information, ranging from interviews to the use of self-reported instruments. In other words, survey research could describe any research design that primarily uses data collected from surveys (Gall, Gall, & Borg, 2007). For war games or any applied research project, any time a participant is asked about their knowledge, attitudes, or behavior then the fundamental principles of survey methods applies.

Survey methods include face-to-face interviews with individuals or focus groups. In these cases, responses are typically collected through observation by the interviewer or other recorder. Survey methods also include the use of self-reported instrument to collect information. To distinguish when an instrument is used, Gall et al. (2007) recommend the phrase *use of questionnaires* instead of the general term survey.

Survey research refers to a study designed to determine the existence of certain characteristics in a population. It is typically carried out as a form of descriptive, quantitative research by asking questions and capturing answers. It involves acquiring information about one or more groups of people, such as their traits, opinions, attitudes, or previous experiences (Leedy & Ormrod, 2010).

Survey Purposes

In general, the purpose of surveys is to produce statistics from information collected by asking people questions (Fowler, 2009). When the primary research design is based on the use of surveys, it is considered survey research. The goal of survey research is to learn about a population, typically by inferring from a sample of information collected on that population (Leedy & Ormrod, 2010). A survey research design provides quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population in order to draw inferences (Creswell, 2014).

War games are not survey research because survey results are not intended to make generalizations about a population of warfighters, planners, or strategists. Instead, survey methods in war games are used, through triangulation with other methods, in order to gain deeper insight into a problem or concept. Survey design can be experimental or descriptive. Experimental and quasi-experimental survey designs are used to compare groups, whether

comparing the same group at different times or by comparing that group to other groups. Descriptive survey design is used to provide information about all the participants in a cohort (Fink, 2003a).

War games are not intended to be predictive. Game results should not imply that the situation that resulted will likely occur, rather should examine results if it occurred (Weiner, 1964). Games consist of players that are experts, that is, people with credible and typically advanced knowledge about a certain matter. This approach is based on purposive sampling rather than random sampling. Purposive sampling is used to produce further insight rather than generalization. "Research based on purposive samples can yield important insights, but the results are not generalizable" (Berman & Wang, 2018, p. 91). The value of surveys is that they can provide systematic and objective information about stakeholder preferences and conditions. Moreover, objective and comprehensive surveys help prevent perceptions from becoming reality (Berman & Wang, 2018).

Games create experiences, but the experiences can only happen in the human brain and its workings are hidden (Schell, 2015). Quantification is troublesome in model building, because some variables are either too difficult to quantify, are simply not calculable, or lack a scale of measurement (Quade, 1964). Polski (2019) explains how games may be helpful to address wicked challenges, whereby "war fighting involves behavioral challenges that cannot be solved or addressed by science or probabilistic mathematics alone" (p. 4).

"A war game is generally a source of opinions, attitudes, and judgments about the particular analysis being undertaken, as well as a very rich source of data" (Weiner, 1964, p. 226). For example, it could provide an analysis of some particular aspect independent of the overall evaluation of the game moves, as well as an analysis of the factors that influence the effectiveness of the system, such as political constraints, intelligence requirements, and weather.

Analysis of military operations involves complex adaptive systems that are multiscale and nonlinear, difficult to determine causation despite apparent correlation. Attempts to measure strategic interaction illustrate this challenge, which involves analysis of own actions as a function of own decisions based on the perception of adversary actions (Polski, 2019).

Survey Data and Collection Types

Survey methods can collect information in the form of quantitative and qualitative data. While quantitative data help address problems from a deductive approach, qualitative data help address problems from an inductive approach. While quantitative instruments provide the option of being

anonymous, qualitative face-to-face interviews are more informal and help clarify ambiguous answers that might be captured using an instrument (Leedy & Ormrod, 2010).

Creswell (2014) describes various types of mixed methods designs that combine the use of both quantitative and qualitive data, such as convergent parallel, explanatory sequential, exploratory sequential, and embedded mixed methods designs. Convergent parallel designs concern merging two databases to show how the data converge or diverge in order to compare different perspectives drawn from quantitative and qualitative data. Explanatory sequential design collects quantitative data followed by qualitative data to help explain the quantitative results. Exploratory sequential design collects qualitative data followed by quantitative data to explore the prevalence of the qualitative insights. Embedded design involves understanding of participant views within the context of an experiential intervention in order to understand experiential results by incorporating perspectives of individuals. Embedded design could be nested within larger designs, such as convergent design.

Types of collection methods include self-administered questionnaires, interviews, and structured observations (Fink, 2003a). Survey data can be collected in a cross-sectional approach, meaning data are collected at one point in time. Also, it can be collected over time representing a longitudinal approach (Creswell, 2014). The three keys to designing good survey questions involve relating them to the research objectives, testing, and putting them in a form to ease collection and analysis (Fowler, 2009).

According to Litwin (2003), psychometrics, as a branch of survey research, enable a researcher to determine how good their survey is. As a field of measurement psychology, it is a way to quantify the precision of the measurement of qualitative concepts. According to Fink (2003a), a researcher collects qualitative data when the literature fails to guide one in designing closed questions or if one wants detailed information in the respondents' words. Qualitative data might be more suitable when there is a small sample of respondents, whereby quantitative analysis might not produce useful results. Overall, qualitative data are more useful to explore knowledge, feelings, opinions and values of respondents, especially if they are considered experts.

According to de Leeuw (2008), the decision to use interviews versus self-administered questionnaires rests on the level of flexibility required. Self-administered questionnaires are the least flexible, because they rely on a fixed structure. Interviews are the most flexible, because they allow for probing. However, you might want self-administered questionnaires if the topic calls for information that the respondent might want to keep confidential, that is, does not want others to know they said it. Although interviews provide better flexibility, they require a well-trained interviewer and might be inappropriate if probing about a sensitive topic that the respondent might want to avoid. Given the advantages and disadvantages of each collection

approach, a mixed mode survey approach might be able to counterbalance the disadvantages of each.

Practical Example 1.1 – *Global Shipping Game and the Ratification of UNCLOS*

In September 2010, the CNO directed the U.S. Naval War College to study the impact on global shipping patterns as a result of the future opening of the Arctic due to climate change and the Panama Canal expansion. The CNO tasking called for a high level game with results reported back to him by the end of the year. With a short turnaround, the development and analysis of game required a simple game design. The game sought to test assumptions and identify challenges with the protection of global shipping. The strength of the game design was the high level experts from government, industry, and academia representing a variety of perspectives, including industry, infrastructure, insurance, and international law. There were two separate scenarios played simultaneously. Two game cells of experts explored the widening of the Panama Canal in 2020. Two other game cells explored the increased access to commercial shipping through the Arctic passages by 2035. The players were high-level experts; therefore it was not desirable for them to spend lots of their time completing survey instruments, especially typing responses to open-ended questions. Most of the data collected was through observations with recorders writing down the summary of group discussions. With a vast amount of qualitative data that the game would produce, it was essential to collect quantitative data using Likert scale instruments that tested the assumptions of the game. Results from survey questions provided quantitative data to conduct statistical tests on these assumptions. For example, a key assumption in the first move of the game was that the United Nations Convention on the Law of the Sea (UNCLOS) would continue to be non-ratified by the U.S. Senate. After group discussion, the players completed survey instruments with a question that asked them to provide their level on agreement with the statement: Non-ratification of UNCLOS would have a negative impact on the U.S. ability to protect global shipping routes in the future. The survey results indicated players agreed with the statement and there were no statistically significant differences across the player cells. Moreover, the survey results showed a disagreement among player responses during move 2 when players responded to the opposite statement concerning the ratification of UNCLOS having a negative impact on U.S. ability. The survey results were critical in reporting the findings of the game. Subsequently, the official Navy JAG position on ratification of UNCLOS differed from the game results. However, the game intent was not to generalize the knowledge of key assumptions, such as UNCLOS, rather to gain insight on the perceptions from experts across various constituencies concerning the assumptions. Without the survey results using closed fixed-choice answers, making sense of the qualitative data would have taken longer and the interpretation of the qualitative may have seemed anecdotal and not representing the majority of the players. However, the quantitative data from the survey results helped expedite the post-game analysis of the qualitative data, as well as provide context to the other data through the use of triangulation of methods.

Steps for Survey Methods

Whether conducting survey research or using survey methods as part of a study, a typical survey design involves posing questions, collecting responses, summarizing with descriptive statistics, and drawing inferences (Leedy & Ormrod, 2010). Generally, the overall task involves turning a list of questions into a questionnaire (Dillman, 2008). While survey use is commonplace, effective survey design makes critical demands on the researcher (Leedy & Ormrod, 2010). Fink (2003b) offers features of a well-designed survey study as follows:

- Measurable objectives to ensure each concept is clearly defined;
- Appropriate design for the use of interviews, observations, or instruments as part of a cross-sectional or experimental approach (cross-sectional, also known as descriptive or observational, provides one group's opinion or perspective at a particular time);
- Reliable and valid instruments that relate to the research objectives;
- Appropriate analysis; and
- Accurate reporting

Most experts in survey methods provide a similar workflow of activities to develop surveys. However, when conducting survey research to generalize characteristics of a population, many of the activities involve the selection of a random sample. When utilizing survey methods using a non-randomized sample of participants, then the sampling guidance can be ignored. Fowler (2009) provides detailed steps for designing good surveys:

- Define objectives – based on the analysis plan that provides a list of variables.
- Draft questions – these could originate from the use of focus groups for brainstorming or standard questions available from other research. Decide whether to use a single question per item or multiple questions that form an index.
- Pre-survey evaluation – testing the content, format, layout of the instrument, and any skip logic.
- Pre-testing – feedback, debugging, and the time it takes to complete the survey.

The define objectives step is based on the analysis plan that provides a list of variables to be considered when developing survey questions. The draft questions step could utilize questions that originate from the use of a focus group for brainstorming or come from standard questions available from other research projects. When drafting questions, the researcher must decide whether to use a single question per item or create multiple questions that form an index variable representing one item. The pre-survey evaluation step involves testing the content, format, and

layout of the instrument, as well as any skip logic. The pre-testing step includes getting feedback for the purpose of debugging and checking how long it takes to complete the survey instrument.

Fink (2003a) provides expanded detailed steps (Figure 1) for the design of surveys and analysis of results, ranging from identifying the survey objectives to reporting.

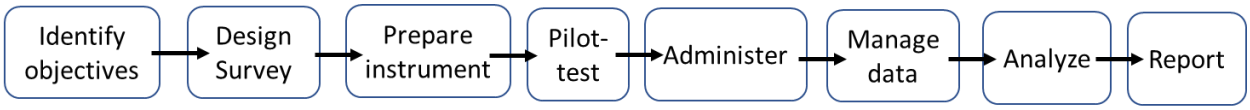

Figure 1. Depiction of survey development steps Adapted from Fink (2003a).

During the identify objectives step, the researcher conducts a literature review to define the terms and underlying theory of the topic. During the design survey step, the researcher determines the type of instruments desired to collect data. During the prepare instrument step, the researcher drafts the questions based on the literature review. The instrument is revised during the pilot-test step. During the administer step, the researcher will conduct training to the data collectors and monitor. During the manage data step, the researcher codes responses, creates a codebook, and cleans data. During the reporting step, the researcher uses lists, charts, and tables to display the results.

Reliability and Validity Considerations

Polski (2019) proffers that rigor consists of validity (is this the right method?), reliability (is the method done right?), and replicability (is it executable?). Regardless of research methods, the goal of quality applied research is to reduce bias and minimize error. For survey methods used in war games, the reliability and validity of the instruments and data collected must be considered. Games produce findings not available by other methods. Therefore, validity in this context evolves from findings that must be accepted as sufficiently valid for the complex situations the games study (Weiner, 1964).

For data collected using survey methods during war games, bias and error can emerge in numerous ways. “In surveys, answers are of interest not intrinsically but because of their relationship to something they are supposed to measure” (Fowler, 2009, p. 87). Information collected is based on self-reported data and pose a risk of the respondent providing a distortion of reality, either on purpose or inadvertent. Some respondents may provide intentional misrepresentations of the facts for various reasons. Other respondents might not have thought about a certain issue until asked or attitudes might be disproportionately influenced by recent events (Leedy & Ormrod, 2010).

Litwin (2003) contends that it is difficult to assess the quality of the data collected, rather it is easier to assess the accuracy of the instrument used as a measure of reliability and validity. Reliability is defined as a statistical measure of reproducibility or stability of data collected using a survey instrument. Reliability has two components in survey research – random error and measurement error. Random error is the unpredictable error that occurs in all research. Researchers deal with this error by establishing a set limit for the amount of random error allowable to reject a null hypothesis. Measurement error deals with how well an instrument performs, assuming that no instrument is perfect. To illustrate, Myers-Briggs is considered as good or better than other personality instruments. On retest, people score in the same 3-4 type preferences at a 75-90% rate. Fink (2003a) explains that survey methods that use a nonprobability sampling approach are convenient and economical but is susceptible to selection bias. This type of sampling is typically used for hard-to-identify groups and for pilot situations. Due to war games using nonprobability sampling, it can be assumed that random error exists and therefore results should not be used to generalize to the entire population that is represented in the war game.

Validity can be defined as the degree to which a survey instrument accurately measures what it purports to measure (Fink, 2003a). In other words, validity is a term used to describe the relationship between an answer and some measure of the true score. Differences between the answer and true score could result from errors in sampling, bias, or scales used (Fowler, 2009). It includes content validity, which refers to how thoroughly and appropriately it assesses the concept; face validity, which refers to how it measures on the surface; criterion validity which refers to the use of a validated survey that already exists; and construct validity which is established experimentally to distinguish groups. External validity refers to how results apply to the survey population (Fink, 2003b). Internal validity refers to instruments that are dependable, that is free of error or bias. It is fair to assume that if it is based on a nonrandom sample, then it is susceptible to selection and membership bias (Fink, 2003b).

Litwin (2003) describes validity as how well the items, scales, and the whole survey measure what its intended to measure, including:

- Face validity – a cursory review of the iems by untrained reviewers;
- Content validity – reviewers who have some knowledge of the subject matter;
- Criterion validity – how well an instrument performs relative to other instruments;
- Concurrent validity – use of other method that is a gold standard to validate instrument;
- Predictive validity – how useful is instrument in forecasting (Ex. SAT scores predict student success);

- Construct validity – how practical is the instrument, which may take years to determine.

Another way to think of construct validity is it tests whether items measure hypothetical constructs or concepts (Creswell, 2014).

Reliability also concerns measures of internal consistency (Creswell, 2014). Internal consistency reliability refers to a group of survey items that measure different aspects of the same concept based on the accepted practice that data are richer and more reliable if several items are used (Litwin, 2003). The concept of determining internal consistency reliability, through the calculation of a Cronbach's Alpha metric, will be discussed in more detail in the Part 2 section during the analysis phase.

Ensuring the rigor in survey methods can be resource intensive and time-consuming. Reducing measurement error through better question design and testing is the least costly way to increase survey value (Fowler, 2009). A game analyst must not create a survey instrument so complex and lengthy that it takes away from the gaming experience. Keep in mind, the goal is to balance the accuracy of the instrument with the ease of completion (Dillman, 2008).

Part 2 – Alignment of Survey Development with the Game Management Process

Tasking Phase

Key survey development activities:

- Draft game objectives
- Conduct literature review

Draft game objectives. During the tasking phase, the game analyst assists with formulating the game objectives that address the problem statement, purpose, and central issue concerning the sponsor (Burns et al, 2013). From the game objectives, the game analyst starts to consider what aspects of the game are either experiential or analytical, what research questions apply to analytical aspects, and whether it requires an inductive or deductive approach.

SMART criteria include *achievable* as a dimension that should be considered when drafting game objectives. To evaluate whether an objective is achievable, the game analyst might contemplate what data can be collected to meet the objective, specifically considering the types of data needed and how best to collect it. Given the data types and collection considerations, the game analyst starts to evaluate the appropriate tools associated with survey methods.

Conduct literature review. During the literature review, the game analyst needs to define all the potentially imprecise or ambiguous terms in the objectives for planning a survey (Fink, 2003a). Understanding the concepts that are central to game objectives provide the starting point for developing a data collection and analysis plan. In addition to researching the game concepts, an initial review of literature might provide examples of survey instruments that might provide useful to the current game.

Instrument development could be based on three approaches: (1) use an intact instrument developed by someone else; (2) use a modified instrument; and (3) use an instrument designed specifically for current research (Creswell, 2014). Krosnick and Presser (2010) recommend reviewing earlier surveys before writing your own for three reasons. First, it is a matter of efficiency. Second, it is a matter of expertise since there are many skillful artisans that have drafted and tested reliable instruments in the past. “Design of questions and questionnaires is an art as well as a science” (p. 299). Finally, earlier surveys may offer analysis opportunities to compare results.

Practical Example 2.1 – *Creating an In-house Alumni Survey*

An alumni relations office at a large public research university grew frustrated with outsourcing survey research of its alumni. The response rates were low at around 1-2% and the information collected was not particularly useful. Therefore, an in-house option to collect survey data from alumni was considered. The first step was to conduct a literature review of alumni survey research efforts at other institutions. The results from the literature review yielded three main types of data to collect from alumni: post-graduation outcomes, satisfaction rates, and marketing information. First, post-graduation outcomes helped assess the effectiveness of the program of study in preparing the graduate for success, such as attending graduate school or gaining employment. Second, satisfaction rates measured how the alumni perceived the quality of programs while they were enrolled. Finally, marketing information concerned ways that alumni stayed engaged with their alma mater. These three areas for data collection served as objectives for the survey research. As survey instruments used at other institutions were reviewed, they could easily be placed in one of these three categories of survey research objectives. A question bank of prospective survey items were developed and prioritized that formed a foundation for survey instrument development going forward.
This practical example demonstrates that good survey research does not start with drafting an instrument, rather defining good objectives grounded in an effective literature review comprises the initial phase of survey development.

Design Phase

Key survey development activities:

- Determine type of game: educational/analytic
- Determine analysis approach: inductive/deductive
- Draft research questions
- Draft data collection and analysis plan (DCAP)

Determine type of game. The type of game designed derives from the overall purpose for the game. Educational games are designed for the players to learn from the experience of the game. Instead of research questions, a game analyst may want to identify what the specific learning outcomes are intended for an educational game. For analytic games, the game will be designed to gain insights on a concept, plan, or capability. A game analyst would identify specific research questions, derived from game objectives, for an analytic game. Most games have components of both educational and analytic purposes (Burns et al, 2013). However, the survey methods employed during the game mainly pertain to the analytic component and it is this area that game analysts focus their attention.

Even for educational games, survey methods can be employed. Data can be collected to help assess what learning the players experienced during the game. There are various ways to conduct learning outcomes assessment in this way, including a pre-test/post-test of knowledge, a summative evaluation of direct measures of learning, or a post-game evaluation survey. Regardless the approach to assess learning outcomes, quantitative data can be collected using some form of survey methods.

Determine analysis approach. For analytic games, the game analyst must consider whether game objectives call for either a deductive or inductive approach to analysis. Again, game objectives may call for a mix of deductive and inductive approaches. Deductive reasoning typically employs aspects of quantitative research (Creswell, 2014). If the game objectives require a deductive approach, then the game analyst must start to think what types of quantitative data could be collected, especially through the use of a closed fixed-choice survey instrument. Inductive logic typically employs aspects of qualitative research. If game objectives and research questions require an inductive approach to answer them, then the game analyst must consider what types of qualitative data could be collected, especially through the use of open-ended questions with collection methods such as interviews or questionnaires.

Draft research questions. For analytic games, research questions are not merely the game objectives restated as a question but "are derived from the game objectives, and inquire about discrete facets of the broader game objectives" (Burns et al., 2013, p. 16). Emerging during the literature review, research questions help guide the study and focus on the explicit concepts, represented as independent and dependent variables, within the scope of the investigation. Independent variables (x) are explanatory or predictive factors, while dependent variables (y) are responses, outcomes, or results as a function of the independent variables (Fink, 2003a).

While research questions are intended to be more focused than the game objectives, they "are not interview questions, research questions are broader, identifying areas to ask questions about" (Merriam, 2009, p. 60). In this way, research questions link the data collection methods, such as interview and survey questionnaires, to the game objectives (see figure 2.1). By following this flow, the game analyst can ensure the findings from the data analysis are relevant to address the sponsor's problem.

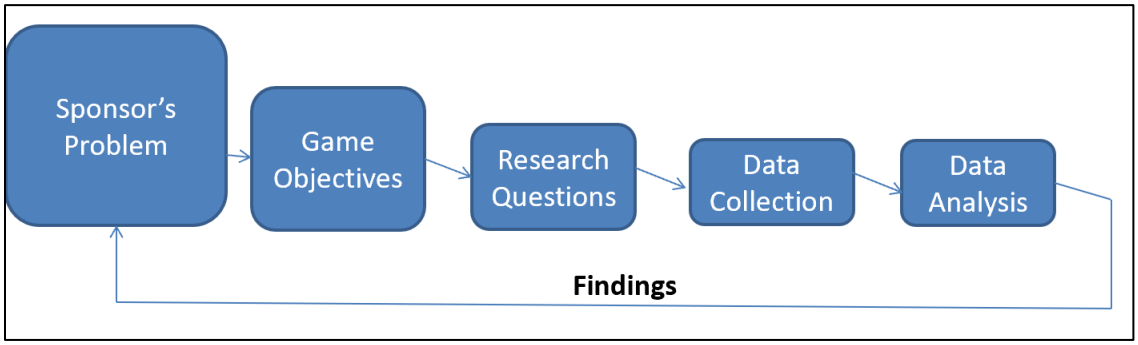

Figure 2.1. Linking findings across research design to address the sponsor's problem.

The game analyst can develop a matrix that relates the research questions to the variables and the survey items (Creswell, 2014). Table 2.1 depicts how to link the research questions, variables, and survey questions using the practical example discussed in practical example 2.1 that refers to creating an in-house alumni survey.

Table 2.1 *Variables, research questions, and items on a survey*

Variable Name	**Research Question**	**Survey Item**
Independent variable 1: Program satisfaction	RQ1: How satisfied was graduate with academic program?	See survey questions 6, 7, and 8: Likert scale for quality of curriculum, quality of instructors, and program relevance to current employment
Dependent variable 1: Annual salary earned	RQ2: What is the total annual compensation earned?	See survey question 22: salary earned from all sources
Control variable 1: First generation status	RQ3: Is the graduate first generation status?	See survey question 28: highest education level for either parent
Relating the independent variable 1 to dependent variable 1	RQ4: (Inferential question) Does program satisfaction influence the annual salary earned?	See survey questions 6, 7, 8 to survey question 22

Note: Table adapted from Creswell (2014)

Draft data collection and analysis plan. Data management concerns the activities to organize survey information to be analyzed. It considers everything from the analysis plan to the start of analysis (Fink, 2003d). The draft analysis plan includes the objectives, research questions, and hypotheses. Table 2.2 depicts how the game type aligns to data collection methods as a function of the analysis framework, approach, and primary data types. Since most games contain multiple objectives that include multiple frameworks, approaches and data types, it can be expected to have multiple collection methods for each game. Ultimately, the key task for the game analyst becomes determining what data collection methods and data types help satisfy the game objectives.

Table 2.2 *Linking game types with data collection methods*

Game Type	Analytic		Experiential
Analysis Framework	Research Questions		Learning Assessment
Approach	Inductive	Deductive	N/A
Primary Data Type	Qualitative	Quantitative	Quantitative
Collection Methods	> Observations of interviews > Open-ended survey items	> Closed fixed-choice survey items	> Pre-test/post-test > Summative direct measures > Post-game evaluation survey

Overall, the DCAP provides guidance for the survey methods that would be appropriate. Fink (2003b) offers advice in choosing a design, such as to consider what is the survey's aim. Is it to describe, compare, or predict? For example, to describe the (concept), to compare (group1) to (group2) or (concept) at different points, or to determine the extent to which (concept x) predicts (concept y). Also, one should determine what are the groups and who are eligible to represent the groups. Finally, one should recognize to what extent the design is experimental or descriptive. Descriptive (commonly referred to as cross-sectional or observational) designs are used to create a portrait of one or more groups at a given time and are more suitable for learning about group rather than a treatment. Therefore, the selection of participants in the group should be random. According to Fink (2003b), survey studies based on nonrandomized control trials are considered quasi-experimental designs. Keep in mind that nonrandomized sampling methods result in

selection bias and membership bias, thus increasing the likelihood of confounding variables influencing the results. However, quasi-experimental designs are intended to measure the impacts of a treatment, as well as being easier to implement (an early example of its use can be found in the bible, Book of Daniel 1: 11-15).

Practical Example 2.2 – *Creating a Survey Model for Measuring Deterrence*

The Deterrence and Escalation Game and Review (DEGRE) series examined the dynamics of strategic deterrence across multiple domains. Survey methods were designed in this game to measure deterrence based on a model derived from a literature review on the concept of deterrence. "Because wargaming alone does not provide answers, the key to assessing deterrence lies with the ability to measure escalation and resolve as the espoused effects and inherent drivers of deterrence decisions" (Ducharme, 2016, p. 45).
The overall purpose of the DEGRE game was analytic and called for a deductive approach. Basically, the research questions asked whether certain actions across domains detered an adversary from escalating the conflict. Strategic deterrence effectiveness proves difficult to measure because it includes concepts such as prevention and manipulation. Using a game theory construct, the strategy employed for strategic deterrence involved brinksmanship, a contest of resolve. Resolve represents a form of risk taking. Therefore, a model was developed for relative resolve as the dependent variable. Relative resolve was a function of the independent variables that included one's own risk tolerance, credible capabilities, and stakes compared to an adversaries' relative resolve.
Given the deductive approach, quantitative data were necessary to analyze the deterrence effectiveness. Closed fixed-choice survey questions were drafted to assess player perceptions of relative resolve and responses were collected using a self-administered questionnaire at the end of each game move. For example, risk tolerance could be measured, using a Likert response scale, by asking about their cell's willingness to escalate as well as the adversary's willingness to escalate. Similar questions were developed to measure capabilities and stakes. The quantitative results showed differences between game cells and changes over the game moves in terms of player perceptions of relative resolve. The results provided context to the actions of the players and helped develop insights on what messaging and signaling were effective for strategic deterrence.

Development Phase

Key survey development activities:

- Refine research questions, develop hypotheses, and align methods
- Create a detailed collection plan
- Draft survey questions

Refine research questions, develop hypotheses, and align methods. During the development phase, the tools and techniques of survey methods are applied. To determine which survey tools and techniques to implement, the game analyst must start with the research questions. Some research questions are worded in an inductive way (see Figure 3). Inductive research questions are more open-ended and drive the collection methods toward survey questions that produce qualitative data. Other research questions are worded in a deductive way. Deductive research questions utilize more fixed-choice data types in its collection approach. Also, a deductive analysis process might call for deriving research hypotheses, including a hull hypothesis. When a research question can be resolved with a distinct answer, then a hypothesis can be stated. A hypothesis is a prediction about how the independent and dependent variables will be related to each other. Statistical tests can be used to analyze this prediction A null hypothesis refers to no relationship exists between variable and will be found in the descriptive statistics through use of a test of statistical significance (Gall et al., 2007). A game analyst must keep in mind that statistical tests should be used to analyze data associated with game objectives requiring a deductive approach.

	Inductive		Deductive	
Game Objectives	To identify… To explore… To analyze… To study… To discover… To report…	To seek to understand… To describe… To report… To uncover…	To assess… To examine… To show relationship… To determine…	To investigate… To compare… To identify differences… To what extent…
Research Questions	How are… How does *x* relate to *y*… What are the differences…		Is there an interaction… Does … explain the relationship between *x* and *y*? Is there a significant difference between *groups* for *x* on *y*? Does *x* relate to *y,* controlling for *z*?	
	To what extent and in what ways do… How do… explain why…			

Figure 3. Common wording for game objectives and research questions.

Using the data collection and analysis plan (DCAP) as a guide, the specific analysis techniques must be considered that will help answer the research questions or test hypotheses. Figure 4 depicts various analysis techniques based on the purpose of the analysis and the data types available from collection. The

analysis techniques used to compare groups (T-test, ANOVA, Chi-square) and identify relationships (correlation) help answer deductive research questions. Analysis techniques using descriptive statistics and coding help answer inductive research questions.

Analysis Technique	Data Type		Purpose
	Ind var (x)	Dep var (y)	
T-test	Nominal (2 groups)	Ordinal	Compare groups
ANOVA	Nominal (>2 groups)	Ordinal	
Chi-square	Nominal	Nominal	
Correlation	Ordinal	Ordinal	Identify relationships
Descriptive stats	Nominal, ordinal, interval, ratio	Nominal, ordinal, interval, ratio	Frequency counts, measures of central tendency (means, medians, range, std dev)
Coding	Narrative text		Content analysis to develop themes

Figure 4. Analysis techniques aligned with data types and analytic purpose.

Create a detailed collection plan. Data are useful for analysis if they are relevant and interpretable. The relevant attribute pertains to data that are related to the research questions. The interpretable attribute pertains to data that can be measured in some way. To make data more relevant, it is a matter of asking the right questions and asking the question in the right way. Drafting good survey questions are discussed further in the next section. To make data more interpretable, the appropriate scale should be used during the collection process.

When collecting survey data using a closed fixed-choice instrument, data types can include nominal, ordinal, interval, and ratio. When collecting survey data using open-ended questions from interviews or questionnaires, data types can include narrative and numerical, as well as nominal categorical, data types depending on how the data are captured. Figure 5 depicts the various response scales to consider for each data type needed for the analysis needed (as depicted in Figure 4).

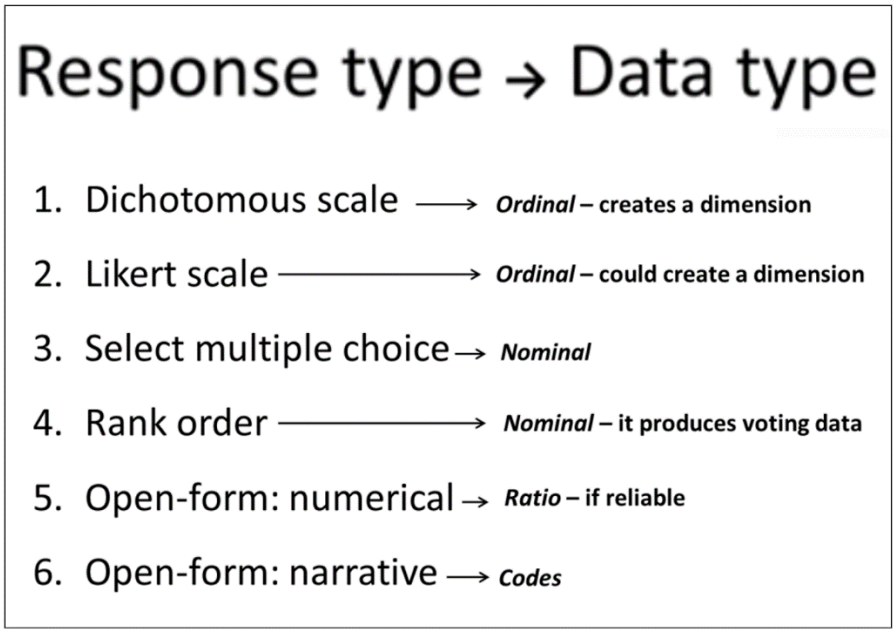

Figure 5. Date type as a function of response scale used.

The goal of the data collection plan is to provide useful data for post-game analysis that answer the research questions. Therefore, relevant and useful data must be obtained for each independent and dependent variables identified from the research questions. Collecting the survey responses using the correct scale provides the appropriate response type. When the data types align with the analysis technique needed to answer a research question, then the data collection aligns to answer the research questions, because the independent and dependent variables comprise both (Figure 6).

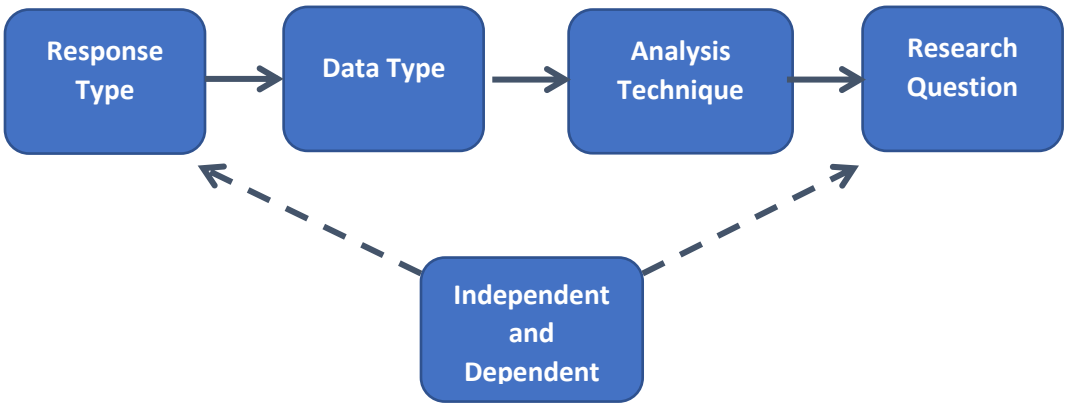

Figure 6. Linkage between response type and research questions.

The next section addresses how to ask the questions to collect the right data. However, the game analyst must consider how the questions will be asked, how the data are captured, and when the questions are asked during the game. Collection methods include interviews or focus groups with observations captured by a data collector, self-administered survey instruments with closed fixed-choice items, self-administered open-ended threaded discussion items, and move templates with questions integrated with the submission. The timing of the survey methods takes into account the context of when the questions are asked during the game. The game analyst must work with the game developer to ensure data are captured at the critical points of the game. In this way, considerations for time spent playing must be balanced with time spent collecting data. Incorporating the survey question items into the move templates are an example of a time-saving activity, such as by using *rock throwing forms* (practical example 2.3).

> **Practical Example 2.3** – *Rock Throwing Forms*
>
> For games with a large order of battle and substantial quantity of planning activities, game developers and analysts found it difficult to incorporate survey instruments and group discussions into the game play. For post-game analysis, the necessary data elements included variables associated with what the actions and missions the players decided to conduct as well as the reasons why the players chose those actions, such as what actions led to a particular action and what effects the player expected would result. Therefore, survey question items were integrated with the move templates, enabling player provided responses for each action using appropriate multiple choice, ranking, and Likert scales. This approach allowed the game analyst to understand the perceptions and desired effects associated with the player actions.

Draft survey questions. After the game analyst works out with the game developer when and how the survey questions will be asked, the actual survey questions need to be drafted. There are two approaches to drafting the initial set of questions. First, during the literature review, previous examples of survey instruments and questions may have provided a good library of questions to use or modify. If previous questions are lacking, then questions need to be created from scratch. The recommendation is to avoid trying to make the questions perfect in the initial draft. Just get questions drafted for each concept and revise later using a good testing protocol. Multiple iterations of survey questions can be expected, because "formulating good questions is an art built on practice" (Berman & Wang, 2018, p. 87).

Start by thinking about how each concept can be quantified (Leedy & Ormrod, 2010). Restrict each question to a single concept. The concepts are derived from the independent and dependent variables identified in the research questions. Appendix A provides tips for writing good survey questions. Fink (2003c) states that the selection and wording of questions are a function of its

purpose, who asks the questions, how asked, who answers, and the characteristics of the respondents. The introduction to the survey should include a purpose of the survey, including elements of the hoped-for outcome, the general topic, definitions used to clarify ambiguous terms, and the specific survey objectives based on the research questions or hypotheses. The researcher must know the culture of the respondents in order to consider whether the questions have appropriate terminology and applicability that improve the ease the completion and accuracy of the data collected.

Using the tips from Appendix A, the creation of a good measure increasing the reliability and validity of the instrument used (Fowler, 2009). However, some concepts are plain difficult to measure. Fink (2003c) provides guidance on collecting information about respondents' attitudes, behavior, and knowledge. Attitudes include opinions, beliefs, preferences, feelings, and values. "Attitudes are very complex entities, and they are difficult to define and measure" (p. 68). In general, attitudes have two components: how they feel and how strongly they feel. Questions about behavior and knowledge are easier to develop. Behavior is basically asking what the respondent actually did. Knowledge is used to assess whether the respondent's opinions are warranted or helps explain their attitudes and behavior. However, asking questions about knowledge could be worded in a threatening way.

Schwarz et al. (2008) described the psychological basis associated with asking various survey question types. Questions about opinions and behaviors have distinct challenges for the respondent. Opinion questions call for not ready-for-use answers that must be formed from judgment based on relevant information that comes to mind. Behavior questions require the respondent to retrieve a recent relevant episode and rely on inference and estimation strategies for the answer. Respondents that must provide judgments on opinions or behaviors can rarely report it in their own minds and need to format it to fit response alternatives provided by the researcher. Thus, numeric scales can provide the solution, as long as it is clear what the numbers on the scales represent. Numeric scales based on behavior amounts (frequency scale) work better than a subjective degree of opinion. If you must ask about opinions, keep in mind that "attitude measurement is highly context dependent" (p. 28). The context of the question could be influenced by the order the question was asked (assimilation effect), the mental construct (formation of standard), contrast effects, cultural differences, or response order (primacy effect meaning first option is most likely).

For each question item, the game analyst must decide whether to use open-ended or closed fixed-choice questions. Open-ended questions tend to be the easiest to draft, test, and revise. It basically entails asking questions and capturing as much qualitative data as possible. The post-

game analysis can be very time-consuming and challenging with qualitative comments. However, while open-ended responses might be difficult to compare and interpret, they provide answers in the respondent's own words, resulting in quotable material (Fink, 2003c). Closed fixed-choice questions are more difficult to prepare but easier to analyze and more reliable (Fink, 2003a), because closed fixed-choice responses provide standardized data that can be analyzed statistically (Fink, 2003c). Though it requires specialized knowledge of statistical techniques to conduct the analysis.

Regardless of question type, one should standardize the response form by providing the context of the question. There are three types of overall contexts for questions: purposeful question that relate to the objectives, concrete questions that are unambiguous, and time period questions (Fink, 2003c). Some questions are like trying to capture a fleeting moment in time (Leedy & Ormrod, 2010). Use open-ended questions to seek clarifying information as a follow-up to closed format questions (Leedy & Ormrod, 2010). Use open-ended questions when intricacies of an issue are still unknown (Fink, 2003a). Krosnick and Presser (2010) offer the pros and cons of using open-ended questions. Its use results in high costs for the analyst, because the analysis depends on grouping responses in a relatively small group of categories requiring the use of a coding scheme, as well as the need for attainment of a high level of agreement between codes. These high costs have facilitated the proliferation of using closed fixed-choice questions. However, an analyst must recognize that "open questions can add richness to survey results that is difficult, if not impossible, to achieve with closed questions" (p. 268). According to Fowler (2009), the use of "why" questions poses problems due to causality considerations. A solution would be to re-word the *why* question as *what characteristics, actions, or signals led you to the decision*?

When considering closed or open form, remember that closed results, such as from using a Likert scale, are easier for quantitative analysis (Gall et al., 2007). For example, when using questionnaires in the measurement of attitudes, one should use a scale to ensure validity and reliability. Concerning self-administered instruments, Fowler (2009) recommends maximizing the use of closed questions for a number of reasons. First, "self-administered open answers often do not produce useful data" (p. 72) due to a lack of probing, responses that are inconsistent, not comparable, difficult to code, or results being perceived as anecdotal. Some respondents prefer to answer closed fixed-choice questions because they are unwilling or unable to express themselves (Fink, 2003a). One should consider that respondent optimizing versus satisfying as a function of task difficulty and motivation (Krosnick & Presser, 2010).

When using closed fixed-choice scales, data types of responses include nominal (categorical), ordinal, and numerical (Fink, 2003a). Numerical response types include interval and ratio scales. A ratio scale is an interval scale that has a true zero. Ordinal can be considered numerical for statistical purposes in some cases (Fink, 2003c). Scales provide data to conduct factor analysis used to determine what are the variables, represented as items, that comprise a measure of that factor or trait (Litwin, 2003). Response options can include a checklist or rating scale, such as a Likert. A Likert scale is typically an agreement statement (Litwin, 2003). Scales are used because behaviors and attitudes are complex and may not easily be quantified (Leedy & Ormrod, 2010).

There are numerous versions of Likert scales to use. Refer to Appendix B to determine which scale is most appropriate for the concept that is being measured. For example, when choosing on a response scale, consideration must be given to the use of a middle point and "do not know" option (Fink, 2003c). Including a *don't know* option in response scales may improve the reliability of the results. The *don't know* option considers the base of familiarity for the respondents. When possible, the use of a screening question should be used instead of the *don't know* option (Fowler, 2009). Also, specialized wording for subgroups might be appropriate to address the different aspects of familiarity across respondents. Due to challenges with using a no opinion option, use of screen out logic is recommended instead (Gall et al., 2007). However, when using skip logic to administer the survey to different groups, it is fraught with risk if not pilot tested (Litwin, 2003).

Krosnick and Presser (2010) reviewed numerous studies conducted to assess rating scales. The number of points on a rating scale varies from 2 to 101. There appears to be no standard and common practices vary widely. The greater the number of points, respondents must have relatively precise and stable understanding of the meaning and interpretation of each point on the scale. The easiest scale for a respondent is dichotomous, such as "like" or "dislike." One approach to try is the use of labels instead of numbers for the scale, because it may be easier for respondents to make mental representation of the meaning and interpretation. The use of mid-point can be problematic depending on the context of the question. The danger of use a mid-point concerns the effect of satisfying by the respondent, especially if they have low motivation to respond. Also, there is considerable evidence that offering a *don't know* option does not improve measurement, because it discourages people to express meaningful answers.

Practical Example 2.4 – *Creating an Index Variable*

Sometimes a concept proves difficult to measure accurately with one question. Therefore, it might be better to create a dimension for the concept that uses multiple questions. The technique to do this involves creating an index variable.
For example, a researcher wanted to measure bias concerning the use of qualitative methods in institutional research (Ducharme, 2014). Bias is a difficult attitude to quantify. But after a thorough literature review, the following factors were cited as possible causes of bias concerning the use of qualitative methods.
Qualitative methods:

- Produce findings that are easy to understand
- Complement quantitative methods
- Are the most appropriate tools for some IR functions
- Increase the credibility of IR findings
- Are rigorous and scientific
- Are more difficult to use than quantitative methods
- Should be used more in IR functions

A 5-point Likert scale was used to capture responses ranging from strongly agree to strongly disagree to the causes of bias. The results were turned into an index variable, However, the Cronbach's Alpha test for internal reliability revealed a score of 0.64. If this measure were above 0.80 (some researchers use 0.70 as a minimum threshold), then the index variable for qualitative method bias could be used for other statistical tests comparing it to qualitative use. However, since the Cronbach's Alpha fell below the threshold, the researcher had to compare individual responses of bias to the use of qualitative methods, resulting in two bias factors that were positively correlated to qualitative method use: easy to understand and complement quantitative methods.

Testing Phase

There is nothing permanent but change. – Greek philosopher Heraclitus

Key survey development activities:

- Pilot testing survey questions
- Revise survey questions
- Analysis preparation

Pilot testing survey questions. Keep in mind that "even carefully designed questionnaires may contain errors and that a questionnaire should always be evaluated and pretested before it may be used in a survey" (de Leeuw et al., 2008, p. 11). Campanelli (2008) describes the three stages of testing and their purpose:

- Developmental Stage – to explore subject matter during the literature review, consult experts on the topic, and consider whether respondents will know the terminology and topic;
- Question Testing Stage – to examine individual questions and the flow of the questionnaire;
- Dress Rehearsal Stage – to test the survey under real conditions.

Campanelli (2008) proposes four basic methods while testing to include (1) informal testing, (2) use of experts, (3) cognitive interviews or focus groups, and (4) actual field condition test. It is best to combine methods for testing. For example, methods 2 and 3 could be combined in the interest of time.

Regardless of the process followed, testing survey questions are a critical activity in the development process. There are two basic approaches to testing: a test for content validity and a test for face validity. The test for content validity involves the use of experts to determine if the right questions are being asked, usually conducted during the alpha test. The game analyst must take into account, given the scenario and other game elements, whether the questions posed to the players would provide sufficient information to help answer the research questions. The use of experts can help deal with this challenge. There are two types of experts to use: subject-matter experts and process experts. Subject-matter experts are those with advanced knowledge of the concepts identified from the research questions. Process experts are those that possess skills in the field of survey methods. Both kinds of experts can help in advancing the content validity of the survey questions.

The test for face validity usually follows the test for content validity. However, when time is limited for testing, the tests can be combined. Face validity concerns asking the questions the right way. The game analyst should try to enter the survey questions in the software that will actually be used or asked by those that will actually facilitate, usually conducted during the beta test. Participants for this test should have similar attributes as the game players. This test gives the game analyst an idea if the questions are fully understandable and being interpreted correctly.

Overall, pilot testing is used to identify errors, problems, and inform redesign needs. Appendix C provides a checklist for pilot testing.

Revise survey questions. Once a testing activity is completed, the game analyst must consider how to revise the survey questions. A good tip for testing is to ask the volunteers in their own words to explain what the question means (Fink, 2003c). When pilot-testing the questionnaire, allow for criticisms and recommendations, then revise and retest (Gall, Gall, & Borg, 2007). Questions should also be unambiguous, specific, and answerable (Berman & Wang,

2018). It may take multiple rounds of testing using various approaches, experts, and volunteers. How many iterations of testing should be done? A game analyst can never be too sure when testing is complete and the survey questions are perfect, but should strive to develop acceptable ones. Campanelli (2008) contends that no expert can write a perfect questionnaire and testing is the only way to assure questions are written as intended. "Remember that any form of question testing is better than none at all" (p. 198).

Analysis preparation. A game analyst should start thinking about how the responses to the survey questions will be analyzed. For open-ended responses, the testing activities provide a good opportunity to brainstorm a coding list to conduct qualitative analysis. For closed fixed-choice responses, consider whether the information would be sufficient to address each hypothesis. Campanelli (2008) suggests a researcher should consider how the data will be analyzed after it is collected. For example, test how to actually key in the data and examine it using statistical software.

The game analyst must scrutinize the final product to ensure right questions are being asked. Consider using a two-column table with each question item in one column and the justification of why it is needed in the other column. (table 9.1 in Leedy & Ormrod provides a good example of this technique). Keep in mind that there could be tendency to collect data solely to support a predetermined solution. However, biased questions are a threat to validity (Berman & Wang, 2018). Therefore, the game analyst must scrutinize the final product to ensure the questions are asked the right way.

Rehearsal Phase

Key survey development activities:

- Train data collectors

Train data collectors. From a survey development perspective, the rehearsal can be considered a continuation of final testing coupled with an execution walk-through. It is not inconceivable that revisions are made to survey instruments and interview questions during the rehearsal. However, major revisions should not be expected as this point. For survey instruments that will collect self-administered responses, proper operation of the survey software is checked. For self-administered move templates, proper operation of the move tool is checked. For questions asked by an interviewer or facilitator, proper techniques for asking questions as well as techniques to collect the observed responses are checked. When collecting survey data through the use of interviews and observations, the quality of this approach depends on training,

standardization, probing, and recording (Fowler, 2009). Keep in mind, the interviewer or facilitator is a data collector and just as important as the observer capturing the responses. To reduce bias, it is just as important for how the questions are asked as to how the observed responses are written down. Refer to Appendix D for procedures to train data collectors.

Execution Phase

Key survey development activities:

- Monitor data collection

Monitor data collection. During the game, the game analyst ensures the facilitators are asking the questions, the observers are collecting the responses, and players are completing any survey instruments. For some games, the same question is asked across different cells and the responses are compared during post-game analysis. For this case, it is important that the facilitators are asking the question in the same way across the cells to ensure standardization. In certain games, the questions asked are very exploratory and facilitators need to be making good use of probing techniques. Experienced facilitators are skilled in this technique, but a quick review of data collected can assist the facilitator in probing in the right areas.

Another method to control for bias is to ensure there is free and open discussion among the players. To speak freely, players need to trust that their responses will not be attributed to them in reporting. Data collectors must adhere to ethical standards in order to protect human rights to ensure the nondisclosure of participant responses (Fowler, 2009). The non-attribution policy at the U.S. Naval War College represents a mechanism to satisfy the protection of human rights.

Ensuring the confidentially of responses could provide one way to increase response rate. However, the game analyst needs to assess whether survey instruments are being completed as expected. Reminders to facilitators and players may need to be exercised. The best technique to increase response rates is to demonstrate the importance of the survey information for the purpose of the research. Best practices, such as providing in-game summaries of select survey results, can motivate players to complete surveys.

Sometimes, in-game revisions to survey questions, either through interviews or on instruments, might be necessary to fix a mistake, clarify a concept, or probe deeper into a concept that emerged during the game. Often, the questions to be used in the final plenary might not be drafted until after a few days of game play occurs. There have been times when final survey instruments are not completed until the initial results from game play informs the questions to be used. The Navy-Marine Corps Command Relationship Game (practical example

2.5) illustrates how an in-process survey instrument could be developed. It is a good example of using a focus group format to develop survey content and format (Fink, 2003c).

Practical Example 2.5 – *Navy-Marine Corps Command Relationship Game*

It is possible to draft the survey questions during the game. This was the case for the Nany-Marine Corps Command Relational Game in 2009. We suspected there might be bias between Navy and Marine Corps leaders on the OPCON and TACON authorities needed by the fleet commander for embarked Marine forces. Therefore, an analysis technique called Analysis for Competing Hypotheses (ACH) was used to design a collection plan that analyzed various command relationships of embarked Marine forces. During game play, the data collection team identified and collected various problems with the command relationships discussed in the game. After each move, those problems were entered as statements in the survey tool and the control cell conducted a rapid survey test just before the final plenary. After the final plenary, the players took the survey instrument and evaluated those problems using the ACH construct. Having quantitative data for the post-game analysis was quite helpful. Reading the qualitative results from the game suggested there were problems with all the command relationships, but it was the survey data that helped identify that one command relationship was preferred over the others and the qualitative data provided the context as to why. By using the ACH construct to minimize bias in research, this game leveraged focus group activities to create the survey questions that were used to evaluate the insights of the game after its conclusion.

Analysis Phase

Key survey development activities:

- Aggregate standardized data
- Conduct data analysis
- Report findings

Aggregate standardized data. Upon game completion, the process of aggregating the data collected into a standardized format involves preparing it for analysis. Preparing data for analysis includes formatting, coding, and cleaning. Formatting may involve turning quantitative data into a flat file for statistical analysis. Coding of qualitative data may require developing a codebook. Cleaning may require addressing nonresponse corrections (Fowler, 2009). After entering the data into appropriate statistical software (SPSS, SAS, STATA, Excel), the function of cleaning data comprises of running frequency tables to verify an miscoding, missing data, and incorrect entries (Fink, 2003d). Most statistical tests are conducted using the data represented as a flat file, meaning one row per survey respondent.

Early in the data analysis phase, it is important to identify missing data, examine its causes, and decide how to handle it (Fink, 2003d). A codebook can be created to facilitate making subjective decisions that arise, such as how to deal with missing items in quantitative data (Litwin, 2003). A codebook provides a description of the variable names, written such that future analysts could derive the same results with the same data (Fink, 2003d). Using a codebook is very important if multiple people are coding. Intercoder reliability can be calculated using a kappa measure.

Conduct data analysis. Various research designs dictate varied approaches to conduct the data analysis. Fink (2003a) suggests the goal of survey analysis is to describe, summarize, compare, and predict based on the data collected. It starts with a frequency count of responses. It continues with comparisons to determine statistically significance differences among groups. It attempts to predict through the use of correlation. If both quantitative and qualitative data have been collected, then a good practice is to start with the quantitative data analysis to identify differences between groups and relationships across variables in order to inform the subsequent qualitative coding. Gall, Gall, & Borg (2007) provide guidance for analyzing questionnaire data based on qualitative techniques or quantitative statistical techniques. The statistical techniques depend on the data types, either continuous or discrete, and include various scales such as interval, ratio, nominal (categorical), or ordinal. The three statistical technique approaches include: (1) measures of central tendency (mean, median, mode, skewness); (2) measures of variability (standard deviation, normal curve); and (3) correlational (bivariate, multivariate). Fink (2003d) provides the following guidance for using statistics:

- Describe respondents,
- Describe responses,
- Identify relationships,
- Conduct comparisons, and
- Test predictions.

According to Creswell (2014), the steps of data analysis and interpretation include:

Step 1. Report descriptive data about the respondents.

Step 2. Determine response bias. (Nonresponse considerations are not applicable for most cases associated with survey methods for gaming research.)

Step 3. Descriptive analysis of x and y variables, including means, standard deviation, and range.

Step 4. Combine items into an index variable (optimal), using Cronbach's Alpha to check internal reliability.

Step 5. Conduct inferential research by using statistical tests to compare groups based on research questions and hypotheses.
Step 6. Conduct interpretation such that the researcher draws conclusions about the larger meaning of the results; present results in tables or figures; and includes tests for statistical significance, confidence intervals, and effect size. Effect size identifies the strength of the conclusion about group differences or relationships among variables in quantitative studies.

There are numerous statistical techniques that a game analyst can employ. However, the three common techniques that are used in post-game analysis include (1) descriptive statistics, (2) causal-comparative analysis, and (3) correlational analysis.

Descriptive statistics. Defined as mathematical techniques for organizing, summarizing, and displaying a set of numerical data (Gall, Gall, & Borg, 2007), descriptive statistics includes frequencies, proportions, means, medians, standard deviations, and other measures of central tendency and variance. Data can be evaluated in either raw or summarized form. Raw data of groups or responses can be described using frequencies and percentiles. Summarized data uses measures of central tendency and distributions to compare groups and responses.

The frequency of responses should be calculated first to identify any missing, neutral, or outlying responses. Missing data due to nonresponse situations are only a problem if there is a difference between responders and nonresponders. To control for this, conduct a statistical nonresponse adjustment after the data are collected by determining the representativeness of the responses through comparing the demographics of the responders and the nonresponders, then applying a weighting scheme to the results as necessary (de Leeuw et al., 2008).

The overall goal for producing descriptive statistics centers on determination of whether the data collected are reliable (consistent) and valid (accurate). Sometimes it is more accurate to use multiple items to measure a concept. This approach involves creating an index variable. To verify if it is appropriate to use an index variable, conduct a test for internal reliability using a Cronbach's Alpha test. Reliable data mean they are consistent and can be determined by evaluating the internal consistency, again by using a Cronback's Alpha test. For data collected by observers, the observation instruments can be evaluated for inter-rater reliability (Fink, 2003a).

Causal-comparative analysis. Defined as an approach that seeks to discover possible causes and effects of a personal characteristic, causal-comparative research compares individuals in whom it is present with individuals in whom it is absent or present to a lesser degree (Gall, Gall, & Borg, 2007). For example, a researcher conducts a factorial experiment to study how two or more independent variables (factors) affect the dependent variable (Gall, Gall, & Borg, 2007). Use of factorial design to present data includes a descriptive factorial approach that uses

crosstabs to present results based on causal comparative testing (Fink, 2003b).The most common and simplest approach to comparative testing is using a *t*-test method. Others statistical tests include Chi-square and ANOVA. Conducting comparisons requires the use statistical significance, meaning that the differences that have been found are statistically meaningful and not explained by chance alone (Fink, 2003d).

Defined as a test of statistical significance that is used to determine whether the null hypothesis that two sample means come from identical populations can be rejected (Gall, Gall, & Borg, 2007), the *t*-test technique uses an independent variable that is a nominal data type comprising two groups and a dependent variable that is continuous such as ordinal, internal, and ratio data types. The steps for conducting a *t*-test can be found online simply by searching for "how to conduct a *t*-test in excel" and numerous instructional links are provided. Once the steps are followed, a *p* value will be produced. If the *p* value is less than 0.05, then the null hypothesis must be rejected. The null hypothesis for any test means that no difference exists between the groups. If it is rejected, then a significant difference exists between the two groups based on the dependent variable measured. While the *p* value indicates whether a significant difference exists between groups, the effect size provides meaning associated with how much that variable explains the difference between the groups. Again, one need only search for "how to calculate effect size for *t*-test" and links that contain the steps to calculate a Cohen's *d* will be displayed. Small effect sizes are defined with a Cohen's *d* at about 0.2, medium effect sizes with a Cohen's *d* at about 0.5, and large effect sizes with a Cohen's *d* at about 0.8.

Correlational analysis. Defined as a type of investigation that seeks to discover the direction and magnitude of the relationship among variables (Gall, Gall, & Borg, 2007), correlational research designs are highly useful in studying social science problems. They allow for analysis of relationships among large groups of variables in a single study, as long as both independent and dependent variables are continuous data types. The relationship is defined through calculation of a Pearson coefficient (*r*), which ranges from -1.0 to 1.0, that that is considered significant with a *p* value less than 0.05. The relationship can be further explained through the calculation of the coefficient of determination (*r*-squared), also called explained variance, which explains the degree that one variable explains the variance in another variable. Table 2.3 depicts the relationship strengths for ranges of *r*-squared measures. Bivariate and multivariate regression is an advanced statistic test to determine how much one variable predicts another variable. However, it has not been particularly useful for data collected from games, because games are not intended to be predictive.

Qualitative analysis. Data analysis in qualitative research involves making sense of the data collected by reducing the data, a process of focusing in on some of the data and disregarding other parts of it, because the text data collected can be so dense and rich that not all of the information can be studied. The goal of reducing data is to categorize the data into a small number of themes, typically 5-7 themes (Creswell, 2014).

Coding is the technique used for qualitative analysis. Coding by hand can be a time-consuming process. Therefore, qualitative software programs help researchers organize, sort, and search for information in large databases. Although there are numerous approaches for coding data, there are two basic approaches to use: selective and open coding. Selective coding uses a predetermined list of codes or concepts. The code list might be developed during the literature review or might emerge from the quantitative analysis results. Open coding involves creating the codes or concepts as they appear during the review of qualitative data.

Triangulation. Defined as the use of multiple data-collection methods, data sources, analysis, or theories as supporting evidence for validity of research findings (Gall, Gall, & Borg, 2007), triangulation is the guiding principle for mixed-method or multiple-method research. Practical example 2.6 presents how survey methods can produce insights through the use of triangulation.

Practical Example 2.6 – *NSSE Attrition Study*

Researchers at a large public institution wanted to learn why students drop out of college. The answer was elusive, because student attrition was a difficult concept to measure due to collection challenges. Many times students would fail to officially check out; that is, they just would not register for the next semester and disappear. Tracking them down for a survey or interview ad hoc would expend a lot of resources. However, researchers figured out a way to identify possible reasons for their departure using the data already available.

The solution was to use NSSE data, which stands the National Survey of Student Engagement. NSSE is a survey tool widely used across almost all undergraduate institutions. It is based on the theory that the more students are engaged in their higher education experience, then better they will develop as a student, such as get better grades, are more likely to be retained, and are more likely to graduate on time. The survey is not intended to be a study of the students, rather it is a good measure to see how well an institution is doing to engage students compared to other similar institutions. Thus, it provides benchmark data for peer institutions to determine what areas they are doing better or worse.

While providing benchmark data is the main use of this survey research, researchers used the survey data to address the problem of understanding student attrition. At typical large public institutions, 20-25% of incoming students never return for the second year. Understanding and intervening the causes of student attrition could have a major impact on revenue as well as student success.

With NSSE data, researchers could go back to all the data collected through NSSE from 2011-2017 and merge it with data on whether the student stayed enrolled or departed. Staying or departing would represent a dichotomous field to create two groups for comparison. *T*-test

statistics were used to determine if there were significant differences in survey responses between the two groups. Then, the open-ended comments collected at the end of each NSSE survey were used to explain why these differences might exist. For example, an open-ended question was asked: what were satisfied or dissatisfied about your experience at UMass Boston? Again, the comments were divided into the two groups: students that departed and students that stayed.
Qualitative coding of the comments produced interesting results. There were not major differences between the retained and departed groups based on academic support, such as student-faculty interaction, advising services, etc… However, co-curricular activities, or the lack there of, was what students that departed cited as something they were dissatisfied about. Therefore, it suggested that students were departing, not because they were failing academically, but because they wanted to attend an institution that provided a better residential campus experience, such as having more opportunities for sports, clubs, fraternities, and social experiences. So that resulted in investments in student services at the large public institution in order to enhance the experience outside the classroom and improve the student attrition problem.

Report findings. When reporting data analysis results, the analyst must balance the amount of detail with the interpretation of the results. Knowing how to strike the right balance is a result of practice and experience. However, the analysis should base the amount of detail and interpretation on the audience of the report. For senior leadership, an executive summary should be heavy on interpretation including just enough detail or facts to support the findings. For the main report, intended primarily for other researchers, it should include copious detail to allow an independent interpretation of the findings as well as elements of the methodology used. "Researchers reporting survey estimates have a scientific obligation to provide a full description of the details of the procedures they used that could affect those estimates. In addition, they should perform and report calculations relevant to the precision and accuracy of their figures" (Fowler, 2009, p. 171).

When reporting on survey results, Fink (2003e) recommends the following elements to be included in the written report:

- Survey method characteristics – type of survey;
- Contents – number of questions and response scales;
- Reliability of measures – if applicable, what index variables used;
- Validity – what testing was done;
- Results relative to research objectives (refer to Appendix E for guidelines in using lists, charts, and tables to report results);
- Conclusions (findings) / Implications (meaning) / Recommendations (next steps).

The actual survey instruments or interview questionnaires as well as any detained statistics should be included in the appendix.

When presenting survey results, Fink (2003e) emphasizes that the audience, whether technical, nontechnical, or mixed, will influence the level of detail in the slides that are used. Refer to Appendix F for guidelines in creating presentation slides. The following elements provide an outline for a presentation of survey results:

- Purpose of the talk;
- Questions answered by the talk;
- Order of the talk;
- Introduction: state problem, state survey setting, why survey used, catch audience attention;
- Overview of survey: characteristics, content, sample question;
- Psychometrics and reliability;
- Design: comparison of groups or experimental;
- Data analysis results:
    - For each table: tell audience title, headings, contents, interpretations;
    - Present figures as necessary.

References

Berman, E., & Wang, X. (2018). *Essential statistics for public managers & policy analysts* (4th ed.). Thousand Oaks, CA: Sage.

Burns, S., DellaVolpe, D., Babb, R., Miller, N., & Muir, G. (2013). *War Gamer's Handbook, A Guide for Professional War Gamers*. Newport, RI: Naval War College.

Campanelli, P. (2008). Testing survey questions. In E. de Leeux, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 176-200). Retrieved from http://joophox.net/papers/SurveyHandbookCRC.pdf

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

de Leeuw, E. D. (2008). Choosing the method of data collection. In E. de Leeux, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 113-135). Retrieved from http://joophox.net/papers/SurveyHandbookCRC.pdf

de Leeuw, E. D., Hox, J. J., & Dillman, D. A. (2008). The cornerstones of survey research. In E. de Leeux, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 1-17). Retrieved from http://joophox.net/papers/SurveyHandbookCRC.pdf

Dillman, D. A. (2008). The logic and psychology of constructing questionnaires. In E. de Leeux, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 161-175). Retrieved from http://joophox.net/papers/SurveyHandbookCRC.pdf

Ducharme, D. R. (2014). *Factors that influence the use of qualitative methods by institutional researchers* (Doctoral dissertation). Retrieved from https://scholarsarchive.jwu.edu/dissertations/AAI3621977

Ducharme, D. R. (2016). Measuring strategic deterrence: A wargaming approach. *Joint Force Quarterly, 82*, 40-46.

Fink, A. (2003a). The survey handbook. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 1). Thousand Oaks, CA: Sage.

Fink, A. (2003b). How to design survey studies. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 6). Thousand Oaks, CA: Sage.

Fink, A. (2003c). How to ask survey questions. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 2). Thousand Oaks, CA: Sage.

Fink, A. (2003d). How to manage, analyze, and interpret survey data. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 9). Thousand Oaks, CA: Sage.

Fink, A. (2003e). How to report on surveys. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 10). Thousand Oaks, CA: Sage.

Fowler Jr., F. J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Pearson.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*(3), 66-70. Retrieved at http://marketing-bulletin.massey.ac.nz

Krosnick, J. A., & Presser, S. (2010). Questions and questionnaire design. In J. Wright & P. Marsden (Eds.), *Handbook of Survey Research* (2nd ed., pp. 263-313). San Diego, CA: Elsevier.

Leedy, P. D., & Ormrod, J. E. (2010). *Practical research: Planning and design* (9th ed.). Upper Saddle River, NJ: Pearson.

Litwin, M. S. (2003). How to assess and interpret survey psychometrics. In A. Fink (Ed.), *The survey kit* (2nd ed., Vol 8). Thousand Oaks, CA: Sage.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Polski, M. M. (2019). Back to basics: Research design for the operational level of war. *Naval War College Review, 72*(3), 62-83. Retrieved from https://www.jstor.org/stable/26654316

Quade, E. S. (1964). Methods and procedures. In E. Quade (Ed.), *Analysis for military decisions* (pp. 149-176). Santa Monica, CA: RAND.

Schell, J. (2015). *The art of game design: A book of lenses* (2nd ed.) Boca Raton, FL: Taylor & Francis.

Schwarz, N., Knauper, B., Oyserman, D., & Stich, C. (2008). The psychology of asking questions. In E. de Leeux, J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology* (pp. 18-34). Retrieved from http://joophox.net/papers/SurveyHandbookCRC.pdf

Weiner, M.G. (1964). Gaming methods and applications. In E. Quade (Ed.), *Analysis for military decisions* (pp. 217-226). Santa Monica, CA: RAND.

Appendix A

Tips for Survey Question Wording

The following best practices for survey question wording are derived from recommendations in Berman & Wang (2018, p. 83), Fink (2003c), Gall et al. (2007, p. 233), Fowler (2009, Chapter 6), Leedy & Ormrod (2010, p. 194), and Krosnick & Presser (2010, p. 264).

Keep it short: Use shorter questions to save time. Open-ended questions are time-consuming. The length of the survey depends on how much time is available for survey completion and consideration of the minimum number of questions needed for accurate measurement. Eliminate extraneous questions by using an item matrix – a three-column table with topic, question, and information collected.

Use simple syntax: Remain succinct by using complete sentences, clear thoughts, and familiar words. Avoid abbreviations, acronyms, slang, colloquial expressions, jargon, and technical expressions.

Eliminate ambiguity: Clearly write in the audience's language, use conventional language to maximize the understanding, and aim for wording that all respondents will interpret in the same way. Avoid using subjective terms such as *several*, *most*, or *usually*.

Standardize format: Strive for wording that is specific and concrete (as opposed to general and abstract), avoid inadequate working, and provide context to the questions asked. Ask about one thing at a time (avoid double-barreled questions). Avoid negative-worded questions.

Minimize bias: Avoid leading or loaded questions that push respondents toward a referred answer. Check for unwarranted assumptions that a respondent should be familiar with topic.

Optimize response options: Responses should be exhaustive and mutually exclusive. Optimize the number of categories in a scale. Use a Likert scale whenever possible. Ask multiple questions per concept in order to create an index variable.

Optimize question order: State the purpose of the survey in the introduction. Early questions should be easy and pleasant to answer and should build rapport between the respondent and the researcher. The next questions should explicitly address the topic of the survey, as it was described in the introduction. Any questions on the same topic should be grouped together. Questions on the same topic should proceed from general to specific. Questions on sensitive topics that might make respondents uncomfortable or controversial should be placed at the end of the questionnaire. Use skip logic to avoid asking questions that do not apply to respondents.

Appendix B

Choosing a Survey Response Scale

Researchers struggle with selecting the appropriate categories for their survey response scales, specifically whether to include neutral and n/a options. This section addresses this survey design decision and offers guidance for choosing these categories when drafting survey questions. For the purposes of this discussion, this section will examine the basic Likert response type with a 5-category scale.

Response Category Options. For 5-category Likert scales, particularly those that range from strongly disagree to strongly agree, there are four basic response category options (see Figure B1 for examples) to choose from:

A. Likert forced choice (no neutral or n/a)
B. Likert with neutral category (no n/a)
C. Likert forced choice with n/a category
D. Likert with neutral and n/a categories

A.	Strongly Disgree ①		Disagree ②		Agree ③	Strongly Agree ④
B.	Strongly Disgree ①	Disagree ②	Neutral ③	Agree ④	Strongly Agree ⑤	
C.	Strongly Disgree ①	Disagree ②	Agree ③	Strongly Agree ④	N/A ◯	
D.	Strongly Disgree ①	Disagree ②	Neutral ③	Agree ④	Strongly Agree ⑤	N/A ◯

Figure B1. Likert examples

All four survey response category options are legitimate and produce data that can be used effectively in statistical analysis. However, the presence or absence of a neutral category can produce distortions in the data obtained (Garland, 1991). A researcher wants the most reliable data possible, so to determine which option to use, a researcher should address the following two questions:

1. Is it reasonable that the respondent is familiar with the topic of the survey item?
2. Is it reasonable to think that the respondent should have an opinion one way or another about the topic of the survey item?

Respondent Familiarity. The first question explores whether to use the n/a category option. Fowler (2009) refers to this category as the "don't know" option. Some researchers caution that when respondents are asked questions about their own lives, feelings, or experiences and a "don't know" option is offered, it allows the respondent to avoid the work required to give the answer. However, other researchers argue that sometimes there are things asked that the respondent legitimately lacks the knowledge about it. The further the topic is from the immediate experience of the respondent, the more plausible that some respondents would have inadequate knowledge or sufficient familiarity to have formed an opinion about the topic.

Another way for the researcher to determine if the n/a option is warranted is to consider whether the survey item concerns a topic that all members of the group are familiar with. If it does, then it is unnecessary to offer the n/a option. However, if it only applies to some of the respondents in the group, then include the n/a option. But when a large number of n/a responses are anticipated, then the researcher should consider why they are asking that question in the first place. In this case, it may be better to consider writing the survey item in a way that everyone in the group could answer.

Respondent Opinion. The second question addresses whether to include a neutral category. The neutral category could be neither agree or disagree, agree or disagree equally, or some other variant of the neutral response. Fink (2003c) suggests only including a neutral category if it is valid. Some researchers suggest that neutral categories provide respondents with an excuse for not answering. However, some respondents resent not having a neutral option. For this reason, a researcher should avoid the risk of getting respondents frustrated and disengaged from the survey. Respondents that are engaged in the survey items provide the most reliable data.

Garland (1991) suggested that "the way in which people will respond to a balanced Likert type scale with a mid-point is content specific" (p. 67). So how does a researcher determine if the neutral category is warranted? The answer is indeed subjective, but a good guideline to use is to consider if the survey item concerns an assessment of the respondent individually (or as part of a group) or an assessment of some other group or topic. For example, say your spouse plays the role of a researcher and surveys your mood for dinner asking "do you want Italian or seafood tonight?" Providing a neutral or no preference response adds no value your spouse's decision-making. Since the question is asking you to assess your individual preference or attitude, then you should be forced to answer one of the two options provided. However, there are times where survey items concern something other than an individual preference or attitude.

For example, as an election approaches, a researcher may survey respondents to select which of two candidates is best qualified. A respondent could legitimately and objectively assess that both candidates are equally qualified. Therefore, a neutral or equally qualified category is warranted. However, if a researcher asks the respondents to select which candidate would they vote for, then that reflects a personal assessment that is normally based on subjective factors rather than objective factors. The researcher is only offering one vote in this survey election, so the respondent must pick one or the other as a measure of their preference. Thus, for two questions about a candidate, a researcher could get two different answers from the same respondent. Which is correct? Again, it depends on what the researcher is trying to measure: perceptions of candidate qualification levels or measures of potential voter preferences.

Response Category Guideline. When drafting your survey items, use the chart in Figure B2 to determine the appropriate Likert scale to use.

1. Is it reasonable that the respondent is familiar with the topic?	2. Is it reasonable to think that the respondent should have an opinion about the topic?	
	Yes	**No**
Yes	**Likert forced choice (A)**	**Likert with neutral (B)**
No	**Likert forced choice with n/a (C)**	**Likert with neutral and n/a (D)**

Figure B2. Guideline for choosing neutral or n/a survey categories

If stuck, do not obsess over the options. It is recommended to go with your best guess and then test for face validity by having test respondents that are similar to the actual respondents that the survey instrument is intended for. Their feedback will help guide your decision in what categories to include. Remember, testing repeatedly is the best way to ensure the validity of the survey instrument and the reliability of the data collected. Also keep in mind that, as Garland (1991) noted, the use of neutral and n/a categories remains largely a factor based on individual researcher preference.

Appendix C

Pilot Testing Checklist

Litwin (2003) provides the following checklist for pilot testing (p. 67). Pilot testing is a necessary and important part of survey development to identify errors, problems, and inform redesign needs.

- Are there any typographical errors or misspelled words in the instrument?
- Do the item numbers make sense?
- Is the type size big enough to be read easily?
- Is the survey in the best language for the respondents?
- Is the vocabulary appropriate for the respondents?
- Is the survey too long?
- Are the styles of the items too monotonous?
- Are there easy questions appropriately mixed in with difficult questions?
- Do the skip patterns follow a logical order?
- Does the survey format flow well?
- Are the items appropriate for the respondents?
- Are the items sensitive to possible cultural barriers?

Appendix D

Data Collector Training Procedures

The following procedures are adapted from interviewer training guidance found in Fowler (2009, p. 133-144). All professional organizations that use survey methods are concerned about data quality and have at least some kind of training for those collecting data, such as facilitators and ethnographers.

Data collectors can be a significant source of survey error or measurement inconsistency. Nearly all reports of reliability ignore effects of the data collection process on the data used for analysis. Training and supervision of the data collectors, those asking the questions and those recording the responses, serve as the controls to improve reliability and minimize bias.

Training methods. The amount of training time has shown to influence the quality of the data collected. Less than one day of training has been considered unsatisfactory, whereby 2-5 days have been considered the norm. There are 6 methods for the training of data collectors:

- Written material – use of a guide (e.g. facilitator guide) helps when there are detailed instructions on data collection that may be too dense for someone to consume in one training session. Also, a structured guide is important when standardization of responses are desired.
- Lectures and presentation – suitable for training of groups, when insufficient time exists to prepare a guide, or the procedures are straightforward.
- Computer-based tutorials – provides valuable training results when the tutorial is designed to be interactive.
- Planned exercises – most efficient method by taking advantage of the testing and rehearsal sessions.
- Practice role-plays – can be incorporated into the rehearsal process or integrated with a lecture or presentation.
- Observation of early interviews – this method provides training as the project proceeds by looking at initial results and providing feedback.

Topics. Regardless of the method for data collection training, topics should include the background for the study, question wording and probing techniques, bias mitigation strategies, and recording procedures.

- Background of the study – provide an overview of the methodology, such as the type of analysis and how the results might be reported, the purpose of the study, including

sponsorship and game objectives. Information on game objectives is helpful in providing respondents with appropriate answers to questions and helping enlist participant cooperation. Finally, the confidentially policy should be addressed, specifically what assurances the facilitator can give the participants that their responses will not be attributed to individuals in reporting.

- Question wording – refer to Appendix A for tips to review with facilitators. Highlight when it is important that facilitators ask questions in a consistent and standardized way.
- Probing – review techniques to clarify answers, such as using a hypothetical (what if…), devil's advocate (opposing view), ideal position (describe perfect situation), and interpretative (get reaction on explanation) techniques.
- Minimize bias – how to use of interpersonal aspects in a unbiased manner in order to minimize inappropriate interviewer feedback and reduce the interviewer's effects on the data. Bracketing is a technique used to identify the perspectives of the facilitator on the topics addressed prior to the data collection in order to identify possible results that may have been influenced through the interview process.
- Recording procedures – collecting data for open-ended questions through observations is a skill that requires preparation and practice, because not everything can be written down, so the goal is to try to find out ahead of time what might be important. Tips include serializing responses in some way to identify respondents, use a double-space between respondent comments, review an initial set of codes to use, and annotate "observer comments" to distinguish from respondent comments.

Appendix E

Written Report Guidelines

The following tips are adapted from Fink (2003e), who suggests that clarity of information is maximized through the use lists, charts, and tables:

When using lists:

- Use only a few words for each item
- Be consistent (parallel form)
- Utilize a space between lines to make it easier to read
- Use bullets
- Use color sparingly
- Limit to 4 items per slide or 8 items per handout

When using tables:

- Use to summarize and compare data
- Determine comparison groups
- Represent groups to compare in the columns
- Show statistical significance (*p* value) in the notes

When using pie charts:

- Use to represent proportions
- Include notes to explain the source
- Limit to 8 slices (group smaller ones as “other”)
- If displayed side by side, enhance larger one to show growth
- Use color blind ballet

When using bar & line charts:

- Use a title
- Explain what x and y axis means
- State source of data in notes
- Compare groups using vertical bars
- If greater than 6 bar sets, use horizonal bars or line graph
- Provide a key or legend
- Unclutter
- When showing results over time, use a line
- Annotate statistical differences

Appendix F

Guidelines for Presentation Slides

The following tips are adapted from Fink (2003e):

- Limit one main concept per slide
- Allow 1-2 minutes per slide for audience digestion (unless similar slides)
- Limit 8 lines of text per slide
- Limit 7 words per line
- Use bullets
- Be consistent (start with noun or verb for each bullet)
- Use fillers (cartoons, pictures) but be wary of distraction
- Use handouts for details
- Use upper and lower case on slides
- Round to nearest whole number or tenth for percentage
- For tables, limit 5 rows by 6 columns
- Clearly label x and y axis for a chart
- Discuss all information presented
- Don’t overwhelm with animation
- Limit 4 colors per slide
- Proofread (e.g. size of font) and use peer review