AF/A5/7
CAPABILITY DEVELOPMENT GUIDEBOOK

Volume 2J

Document Writing Team

Appendix A – Measures

Appendix B – Survey Research

October 2023

Air Force Futures Requirements Oversight Team

AF/A5DR, Pentagon 5C858

PREFACE

This Guidebook is one in a series of AF/A5/7 developed guides that describes the Air Force processes for the development and validation of operational capability requirements in support of overarching Capability Development efforts. These processes comply with the main processes for Requirements via the Joint Capabilities Integration and Development System (JCIDS), Acquisition via the Defense Acquisition System, and Resourcing via the Air Force Strategy, Planning, Programming, Budgeting and Execution. This guidebook contains general information about facilitation concepts, techniques, and tips, and provides best practices for leading, facilitating, and participating in a Document Writing Team.

Document Writing Teams are used to develop Air Force-sponsored requirements documents and provide consistent cross-functional involvement in requirements generation. The intent is to accelerate the documentation process, improve the quality of the documents, and provide an enduring forum for developing, fielding, and sustaining operational systems. As much as possible, core members of the writing team are maintained throughout the process as the capability matures. Individual participants may change but office symbols will provide continuity. The result is an executable, risk-based, fiscally informed requirement that delivers affordable capabilities at optimal cycle time to the warfighter. Understanding the roles and responsibilities of the team facilitator, lead, and members will enable the facilitator to guide and advise the team and meet the objectives of the document writing event.

Appendix A, Measures, describes a four-step measure development process the analyst can use to properly develop measures for a study. The process can be tailored by the analyst depending on the needs and the requirements of the study. The guidebook also provides insights into various data collection and analysis methods. Finally, guidance on interpreting and communicating results is provided to help the analyst construct a credible and defensible assessment of the results.

Appendix B, Survey Research, describes the principles of survey research that ensure questions are both reliable and valid. With expert elicitation being a special form of survey research, this handbook also presents an approach to conducting expert elicitation in operational capability requirements studies.

The AF/A5DY - Office of Aerospace Studies (OAS) has nearly 20 years of experience in supporting organizations across the Department of Defense and Federal government with analysis training, planning, and execution. OAS provides a full spectrum of analytical assistance in planning and conducting Capabilities-Based Assessments, pre-Materiel Development Decision analyses, Analysis of Alternatives, and assists MAJCOM Document Writing Teams for capability development related documents.

There are no restrictions on release or distribution of this guidebook.

Additional guidance and information to supplement this Guidebook is located on the AF/A5DR Requirements Policy & Integration Portal Page:

- Go to https://www.my.af.mil
- Navigate to “Search AF Portal” and enter the keyword “A5DR”.
- Click on “AF/A5DR Requirements Policy & Integration.”

If you have questions regarding specific information within the Volume 2-series Capability Development Guidebook(s), or if you have suggestions for improvements, please contact:

AF/A5DR: Mr. Richard “Bullet” Tobasco, richard.tobasco.2@us.af.mil, (703) 692-4197, DSN 222

Guidebook OPR: Mr. Jeff “Shredder” Hackman, jeffrey.hackman.1@us.af.mil, (703) 692-1087, DSN 222

Document Writing Team OPR: Mr. Jeffrey Stough, jeffrey.stough.1@us.af.mil, DSN 574 / (757) 764-5018

CHANGE SUMMARY

Change Summary	Date
This document captures updated organizations, roles, responsibilities, and DAF guidance and must be reviewed in its entirety. Portions of this guidebook were derived from the Office of Aerospace Studies, High Performance Team (HPT) Facilitation Guidebook, 18 August 2014, which is rescinded and replaced by this AF/A5/7 Capability Development Guidebook, Volume 2J.	August 2023
Admin changes	October 2023

TABLE OF CONTENTS

SECTION 1. INTRODUCTION

Writing a requirements document is a challenge. The competing interests of the requirements community, the acquisition community, and the planning and programming community contribute necessary friction to the process. The subject matter expertise in the areas of doctrine, operations, training, test, intelligence, cyber, and interoperability affect most every program and make it impossible for one person, or even one office to write a document that will survive the Air Force and Joint staffing processes. Forming a Document Writing Team (DWT) to develop Air Force-sponsored capability development documents is the AF solution to provide consistent cross-functional involvement in requirements generation that produces executable, risk-based, fiscally informed requirements that deliver affordable capabilities at optimal cycle time to the warfighter.

1.1. The Document Writing Team. Understanding the roles and responsibilities of the DWT facilitator, lead, and members will help enable the facilitator to guide and advise the DWT and meet the objectives of the document writing event.

1.1.1. DWT Facilitator. The facilitator guides and advises the DWT to ensure it is productive and worthwhile for all team members and helps enable the team to achieve its objectives. It is important to note that the facilitator is not the DWT lead nor a passive observer of the writing event. Furthermore, the facilitator must be engaged throughout the DWT event and keep the team focused on the topic at hand. The facilitator must have an appreciation for the experience that the members bring to the document writing event. The primary responsibilities include the following:

- Preparing the DWT lead for the document writing event and assisting the team lead in identifying and preparing the other team members.
- Guide and advise the team during the document writing event.
- Providing subject matter expertise on the format and content of the subject document and associated guidance.
- Enabling the team to achieve its objectives.

The first duty of the facilitator is to introduce the document writing process as outlined in the applicable capability development guidebooks and references and help team members effectively implement the process. The facilitator guides participants through the event to ensure they are aware of the standards of performance required, can provide useful input to the document, and can deliver a quality product in the time available.

To enable the team to meet its objectives, the facilitator must have general knowledge of the mission area, capability gaps, and other key references pertinent to the mission area of interest. The facilitator prepares for the writing event by reviewing and understanding the relevant background material. In some situations, the facilitator may need to conduct a literature search for other studies that may have been completed in the mission area of interest. In all cases, the facilitator must engage the document sponsor, Air Staff, and other stakeholders to understand their perspectives, issues, and concerns.

Ideally, the facilitator should establish a rapport with the DWT lead well in advance of the writing event. Through this rapport, the facilitator will be better able to assess the needs of the team (e.g., how much have they done, what needs to be done, what is the level of experience, how best to guide them forward) which will help the facilitator and team lead plan the writing event as well as determine the required resources. The needs assessment will also enable the facilitator to develop the facilitation approach that will be used (e.g., number of days for the writing event, tasks that will be completed on each day, working

groups that will be formed). Working with the team lead prior to the event will also allow the facilitator to gain insights into the politics, issues, subject matter, and personalities involved.

If questions remain or additional information is required as the facilitator prepares for an upcoming DWT, ask an experienced OAS advisor to help. They can recommend policy and document resources that can be used to learn more about the process such as the JCIDS manual, CD Guidebooks, and other references.

1.1.2. DWT Lead**.** The team lead is responsible for communicating details of the HPT event (e.g., dates, meeting location, security, visit requests), identifying DWT members, ensuring members have the permission and funding required to attend the event, distributing read ahead material, writing and sending invitations to identified participants, obtaining funding to conduct the writing event, leading the execution of the writing event, and providing support to document DWT outcomes and actions.

The MAJCOM or organization that is sponsoring the DWT designates a team lead. The team lead must be a military member or government civilian (not a contractor). The team lead has overall responsibility for planning and conducting the document writing event and has the final decision on the content of the DWT products.

1.1.3. DWT Members. Each DWT member plays a vital role in the success of the team. Each member is selected for a specific reason and is expected to contribute to meeting the team's objectives. For example, the member selected based on their background in intelligence is expected to address intelligence-related aspects of the study guidance or study plan such as potential scenarios and threats for consideration, scenario and threat selection methodology development, intelligence mission data requirements, and other intelligence support requirements or issues.

The HPT lead and facilitator must define the expected contributions of each member and establish an environment that is conducive to open and non-confrontational discussions that enable each member to be as productive as possible. The team lead and facilitator should strive to make the document writing event a productive and worthwhile experience for all members.

1.1.4. DWT Support. Experience has shown that scheduling the document writing event, consolidating, and distributing read-ahead materials, recording information during the event, and producing and publishing minutes requires assistance from one or more individuals responsible for managing and accomplishing administrative tasks. It is not advisable for the team lead to attempt to lead and provide administrative support. Having one of more individuals charged with handling the administrative details will help alleviate the administrative burden and enable the team lead to focus on the more important task of leading the DWT.

SECTION 2. FACILITATION FUNDAMENTALS

This section introduces five fundamental concepts of facilitation. Although an understanding of these concepts will be helpful to anyone who must facilitate an DWT or other group, these should not be viewed as rules, absolutes, or magic formulas for successful facilitation. Facilitation, and the work of the DWTs themselves, is a highly human endeavor, more art than a science. These five concepts, Stages of Team Development, Leading a Team, Active Listening, Gaining Consensus, and Human Interaction are broadly applicable facilitation tools and knowing when and how to apply the tools—or in some cases *not* to apply them—is the critical part of the art.

2.1. Stages of Team Development. Like other groups, the DWT will progress through stages of development that are necessary and inevitable for the team to mature, accept challenges, plan work, address problems, develop solutions, and deliver results. Tuckman (1965) developed a four-stage group development model that is relevant to understanding and facilitating the DWT. The model is comprised of four development stages: Forming, Storming, Norming, and Performing.

2.1.1. Stage 1 – Forming. This stage is important since it allows team members to get to know one another and understand the details of the task. Individual team members often desire acceptance by others and avoid conflict and controversy. Team members initially focus on gathering information such as the team objectives, organization, tasks, and schedule. Team members may be motivated, but typically behave independently, focus on themselves, and exhibit their best behavior. During this stage, team members are individually pondering questions such as “Why am I here?”, “Who can I work with?”, and “What are we doing?” Clear and strong leadership helps get team members introduced to each other and involved in the effort. The facilitator must be prepared to answer many questions about the purpose and individual roles and responsibilities of the DWT. The theory at this stage applies to a group brought together at or near the start of the DWT. That may be the case, a DWT is likely to have several core people who have been closely working the issue for an extended period, and a few new members who truly do not know each other. Often the MAJCOM and Product Center representatives may have been part of the perquisite studies and other work and have been working the related development planning issues for some time. This makes the facilitator’s job more “interesting” because parts of the group might be in this stage and parts may be beyond.

2.1.2. Stage 2 – Storming. The team members begin voicing their opinions, aligning with others who share similar views, and confronting others with different views. This stage can be contentious and unpleasant for team members who do not handle conflict well. Sometimes, the tension level may rise to a level that causes arguments to occur among team members. If not properly controlled, this stage can be destructive and adversely affect the motivation of the team. In some cases, teams will not develop past this stage. The facilitator must emphasize being patient and allowing others to express their views. Without tolerance and patience, the team will likely fail to accomplish its tasks. Facilitators must help all team members voice their views, and when needed, work to achieve consensus. Compromises may be required to reach consensus and enable the team to progress. The facilitator should be directive with guidance for decision-making and professional behavior. The facilitator should be accessible to help resolve differences and enable the team to evolve to the next stage. In many ways, this is the most trying stage of the process. The facilitator will often feel serious pressures to hurry up and complete this stage, but if the group artificially decrees they are finished storming, the different views and perspectives will often just go into hiding, not really having been resolved. There are several things to avoid at this stage, none of which is necessarily easy:

- Do not assume every issue will reach consensus. Sometimes there are multiple correct opinions to be addressed. That is why there are different perspectives on the team. The logistician and the operator may both be correct even though at odds.
- Do not easily rely upon voting to shut down serious discussions. It may have its place, but usually only as a last resort, and it hardly ever comes without significant delayed pain. It is often better to accept that diverse opinions that are strongly held and supported on both sides by logic and facts. The issue may need to be resolved after the document writing event.
- Neither the loudest voice nor the voice with the highest rank should necessarily rule.

2.1.3. Stage 3 – Norming. At this stage, agreement and consensus form among the team members. They begin to share a common commitment to achieve the goals of the team. This may have required some members to give up their own ideas and agree with others to enable the team to function. Individual roles and responsibilities are clear and accepted. The facilitator focuses on enabling the team to achieve its goals. In this stage, the facilitator needs deduce whether members are giving up their ideas based on logic and facts, or whether they are feeling significantly outnumbered and pressured to stop delaying the process.

2.1.4. Stage 4 – Performing. Here, the team has achieved a high-level of autonomy and requires little direct involvement from the facilitator beyond maintaining progress. The team is knowledgeable, motivated, and competent. Tasks are delegated t and the team handles decision-making. Although disagreements may still arise, they are likely to be resolved by the team.

2.2. Leading a Team.

Before the document writing event, the facilitator should determine the support required by the team lead. This requires discussion between the facilitator and the team lead to ensure they both understand each other's role and the team lead's expectations. Early and frequent communication and trust is critical. It is important to note that the facilitator is not the team lead's backup. The facilitator and team lead have tasks to accomplish with no single right way to divide the tasks.

In some cases, the DWT lead may have little to no experience in leading a DWT. Consequently, the team lead will likely rely heavily on the facilitator to help lead the document writing event. In other cases, the team lead may have DWT experience and rely less on the facilitator. While leading a team can be a daunting task, an understanding of leadership styles and the needs of the team will enable the facilitator to lead an effective event.

The type of leadership style used by the facilitator will depend on the situation. There is no single best style of leadership, but rather the facilitator must be flexible and adapt the leadership style based on the situation. Effective leadership is task-relevant, which means the facilitator and the leader must use a leadership style that is appropriate for the DWT task and needs of the team.

Hersey and Blanchard (1977) describe a situational leadership model that addresses leadership styles and team maturity levels that are relevant to leading and facilitating an HPT. The model is comprised of four situational leadership styles and four maturity levels of a team. The four leadership styles are:

- S1 (Telling/Directing). Characterized by primarily one-way communication in which the leader defines the roles of the team members and provides the what, how, why, when, and where for the tasks that must be accomplished. In other words, the leader takes a directive role and tells the team what to do and how to do it. The focus is on accomplishing the task and less on the relationship with the team. If the leader focuses more on the relationship, the team may become confused.

- S2 (Selling/Coaching). While the leader is still providing direction, two-way communication and support allows the team members to get on board. The leader provides information and direction, but there is more communication with the team. The leader spends time listening, advising, and coaching. Telling the team what to do may demotivate it or lead to resistance, so the leader must sell the way of working a task by explaining or clarifying an approach.
- S3 (Participating/Supporting). The leader focuses more on the relationship with the team and less on direction. The leader supports and works with the team and shares decision-making responsibilities. The leader spends time listening, praising, and making the team feel good when it demonstrates the necessary commitment.
- S4 (Delegating). The leader passes most of the responsibility onto the team. The leader still monitors progress but is less involved in decisions.

Note that S1 (Telling/Directing) and S2 (Selling/Coaching) are more focused on getting the task accomplished while S3 (Participating/Supporting) and S4 (Delegating) are more focused on developing the team members' abilities to work independently.

The appropriate leadership style will depend on the maturity of the team d. Maturity levels are task-specific, which means team members may be generally skilled, confident, and motivated in their jobs, but the maturity of the team may still be low since it is performing a task requiring skills they may not possess. The four levels of maturity are:

- M1 (Unable and insecure). The team members lack the knowledge, skills, or confidence required to accomplish the task. They are also unwilling to take responsibility for the task and often need to be pushed to take the task on.
- M2 (Unable but willing). The team members are willing to work on the task, but they still do not have the skills to complete it successfully.
- M3 (Capable but lack confidence). The team members are ready, experienced, and able to do the task, but lack the confidence or willingness to take on responsibility. They have more skills than the M2 group, but they are still not confident in their abilities.
- M4 (Very capable and confident). The team members are experienced in the task and comfortable with their abilities to do it well. They are able and willing to accomplish the task, and to take responsibility. They have high confidence and strong skills, and they are committed to the task.

Team members should be selected because they are capable and confident within their topical area of expertise. The A6 representative, for example, should have an M4 maturity when talking about the communications aspects of the task, but may be at an M2 maturity level with respect to requirements development. This dichotomy can result in some behavior that the facilitator needs to understand and address accordingly.

Using the situational leadership model, Hersey and Blanchard (1977) map each leadership style to each maturity level as shown in Table 2-1. The facilitator should assess the maturity of the DWT at the beginning of the document writing event to determine the most appropriate leadership style to use. The team members will likely have different levels of experience. This means the team's maturity level at the beginning of the event can range from M1 (unable and insecure) to M4 (very capable and confident).

Table 2-1: Maturity Level and Most Appropriate Leadership Style

Maturity Level	Most Appropriate Leadership Style
M1 (Unable and insecure)	S1 (Telling/Directing)
M2 (Unable, but willing)	S2 (Selling/Coaching)
M3 (Capable, but lack confidence)	S3 (Participating/Supporting)
M4 (Very capable and confident)	S4 (Delegating)

The facilitator should think about what leadership style(s) he or she is most comfortable with personally. For those styles that are less comfortable, the facilitator may first have to learn some new behaviors before using the styles.

The situational leadership styles and maturity levels of a group are related to Tuckman's four stages of group development described in the previous section in Figure 2- 1. Through effective leadership, the facilitator can help enable the team to progress through the stages of development and accomplish its tasks. During a document writing event, the DWT matures by acquiring knowledge and developing ability. As the team matures, the facilitator adapts his or her leadership style to move to higher levels of leadership and ultimately finish with S4 (delegating). The facilitator should understand this relationship and use the most appropriate leadership style for the maturity level and development stage of the team.

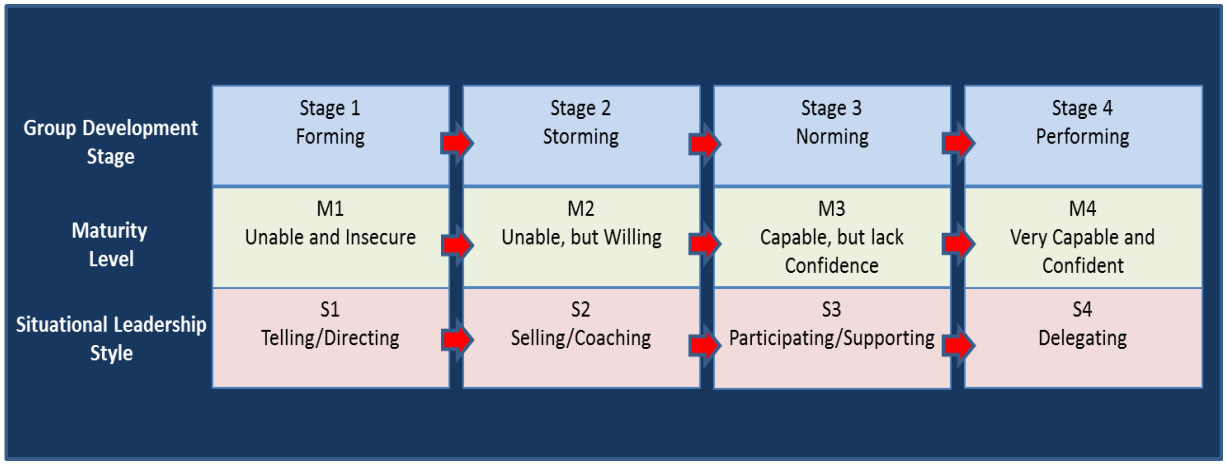


Figure 2-1: Situational Leadership Style, Maturity Level, and Group Development Stage Relationship

As with any interpersonal group effort, the real answer is always "it depends". The discussions above have broad applicability and the work behind these theories applies to many DWTs, but it is not a recipe to be followed by rote, and as mentioned already, the critical skill of the facilitator is to observe, sense, and apply the right tool and approach at the right time. Preexisting relationships among some team members will significantly affect this matrix. How thoroughly the pre-document writing preparation work was, the quality of the read-aheads, and the nature of the problem will all modify the thoughts above.

Regardless of what leadership style is used, the facilitator and team lead are responsible for leading the team and ensuring effective discussion. The facilitator should use the following principles when leading a discussion:

- Foster open discussion. The facilitator should be attentive to the process, content, and interpersonal dynamics of the discussion. The facilitator should ensure no one person or small

group dominates the discussion, and that the discussion is civil and organized. The facilitator should ensure everyone follows the ground rules and that ideas are critically analyzed. The facilitator should establish an environment where minority ideas are voiced freely, and the discussion is respectful.

- Involve all team members. In most teams, one or more members are less assertive, shy, or cannot break in the discussion quickly enough. For these team members, the facilitator should directly ask for their opinions and encourage them with body language or praise. It is important to achieve an exchange of ideas by providing an opportunity for all the team members to participate.
- Ask questions or offer ideas to advance the discussion. The facilitator should be aware of the progress of the discussion and, when necessary, ask questions or provide information to stimulate thinking and move the discussion along. In some situations, the discussion may sidetrack, lose strength, or stall on a problem or "pet rock." It is the facilitator's responsibility to identify the points of agreement or disagreement and ask questions or offer ideas to advance the discussion. Open-ended questions that are not answered with a simple yes or no response are ideal in these situations since they require some thought which could generate discussion. Also, use questions to elicit why someone believes as they do, especially if it seems to go counter to the HPT's "prevailing wisdom". Often the first input is not the real input.
- Summarize important points, arguments, or ideas. Summarizing important ideas, points, or conclusions as raised will ensure that all team members understand what the individual or group meant. y Restating the information and observing the team for verbal or non-verbal clues of understanding may help.
- Make sure that everyone is using the same definition of key words. If someone says, "Everyone knows what *requirements* means," it is likely that there are multiple definitions of a word used by different members, and no, not everyone knows what it means.
- Demonstrate behavior and attitudes. The facilitator should demonstrate the behavior and attitudes desired of the team by:
    - Respecting all team members equally.
    - Being aware of feelings and reactions of team members and responding appropriately.
    - Admitting mistakes or not knowing facts or answers.
    - Being objective and controlling bias.
    - Asking questions based on others' statements.
    - Focusing on positions rather than personalities.
    - Listening carefully and using encouraging body language and tone of voice.
    - Acceding when others have a better idea.
    - Accepting criticism.
    - Providing positive feedback.
    - Giving up the floor when appropriate.
    - Supporting points with fact.

2.3. Active Listening

Listening is one of the most important attributes of a facilitator. Notwithstanding its importance, research indicates that people generally remember only a fraction of what they hear. When communicating, people often wait to speak or become distracted rather than listening attentively. Listening skill can be improved. By becoming a better listener, the facilitator will be able to gain more information and a better understanding of the communication exchange, which will improve his or her ability to facilitate a team.

The way to become a better listener is to practice active listening. Active listening is a communication technique which requires the listener to hear not only the words that another person is saying, but more importantly, try to understand the complete message being sent. The fundamental underpinning of active listening requires the listener to feedback what he or she heard to the speaker by re-stating or paraphrasing the information in his or her own words. In some situations, the listener paraphrases the speaker's words as a question reducing the chances of assumption or interpretation. This not only confirms what the listener heard, but also confirms the understanding between the listener and speaker.

For most people, poor listening habits are often difficult to break. Overcoming poor listening habits typically requires changes in how one comprehends, retains, and responds to the messages received. Hallett (n.d., accessed March 2014), describes five key principles of active listening that, when followed, should help one become a better listener. The facilitator should use these principles to ensure successful communication with HPT members. The five principles follow:

2.3.1. Pay Attention. The listener should give the speaker his or her undivided attention. Key elements of paying attention include the following:

- Look at the speaker directly.
- Put aside distracting thoughts.
- Do not mentally prepare a rebuttal.
- Avoid distraction from environmental factors such as side conversations.
- Read the speaker's body language.

2.3.2. Show Listening. The listener should use his or her own body language and gestures to convey that he or she is listening. Using body language and other signs to indicate listening also reminds the listener to pay attention. The listener should use simple head nods or short affirmative comments which do not necessarily mean agreement, but rather indicate the listener is listening. The listener can show that he or she is listening by the following actions:

- Maintain eye contact.
- Nod occasionally.
- Smile and use other appropriate facial expressions.
- Maintain an open and inviting posture (uncrossed arms, facing position).
- Encourage the speaker to continue with short affirmative comments (e.g., yes, uh huh).

2.3.3. Provide Feedback. The listener's personal filters, assumptions, judgments, and beliefs can distort what is heard. The role of the listener is to understand what is being said, which requires reflecting on what is said. The following tips can help the listener provide feedback to the speaker:

- Paraphrase what is being said into a question, for example:
    - "What I'm hearing is…"

- "Sounds like you are saying…"

- Ask questions to clarify certain points, for example:
  - "What do you mean when you say…"
  - "Is this what you mean…"
- Summarize the speaker's comments periodically.

Oftentimes during a document writing event, DWT members are not listening to the speaker but are instead rehearsing their next statement. One of the most important functions of a facilitator is to provide feedback to both the speaker and member(s) who should be listening using the approaches described above. By doing this, the facilitator redirects attention back to the issue at hand, enabling the HPT to fully address the issue before moving on.

2.3.4. Defer Judgment. Interruptions can frustrate the speaker, be counterproductive, and hinder listeners from gaining a full understanding of the message. Listeners should defer judgment by:

- Allowing the speaker to finish each point before asking questions.
- Avoiding interruptions with counter arguments.

In some situations, the team lead will have the urgency to stay on schedule and may attempt to short circuit the process. An indication that this may be occurring is when the team lead allows one or two people to speak, but then cuts off the conversation prematurely in order to move on. When this happens, the member who was patiently waiting to reply or counter the speaker will now feel shut-out. As a result, the member loses patience and starts interrupting conversations to get his or her perspective heard. Such behavior can be very detrimental to the writing process. It is important to allow all members the opportunity to respond. Nothing is gained by being disrespectful, cutting off, or verbally attacking the speaker. Listeners should respond appropriately by:

- Being candid, open, and honest
- Asserting opinions respectfully
- Being polite to the speaker

2.4. Gaining Consensus. During document writing events, there may be times when the facilitator must gain consensus on a decision to keep the team moving forward. Consensus means overwhelming agreement but does not necessarily mean unanimity (A Short Guide to Building Consensus (n.d., accessed March 2014)). Although the facilitator should aim to achieve unanimity in all decisions, there will likely be situations when one or more team members may be holdouts (i.e., team members who think their interests may be better served by not agreeing with a decision). Interests are the underlying needs or reasons why team members take positions or make demands. In these situations, the facilitator should settle for consensus that goes as far as possible toward meeting the interests of all team members. The following describes a four-step approach for gaining consensus:

- Step 1. Present the Position. The facilitator asks the team member(s) who is holding out to explain why he or she is taking a certain position. To help the team understand, the facilitator can ask the team member to explain his or her position with an example or citation.
- Step 2. Ask Questions. After the team member has expressed his or her position, asking questions helps the team gather more information to understand the underlying reasons why the team member is taking a certain position. The facilitator and other team members should be actively

listening rather than thinking about a response. The facilitator should allow time for the team to consider the new information.

- Step 3. Discuss Modifications. The facilitator should ask the team member(s) who is holding out to suggest modifications to the decision that would make it acceptable to him or her without making it less acceptable to the other team members.
- Step 4. Make Decision. If the modifications are not acceptable to the other team members and unanimity is not achieved, it is appropriate to settle for consensus (overwhelming agreement) on the decision. Great care must be taken not to let the majority, even an "overwhelming majority" override the lone voice if the lone voice is the expert in that area. When deciding how to do brain surgery, fifty patients in overwhelming agreement should not carry the decision over the one brain surgeon who believes otherwise. It is also important to note that the dissenter's opinion will not die in the HPT. The study will have to go through formal coordination and the dissenter's office may make a critical comment on the issue at that time, which may need to be resolved at a higher level.

2.5. Human Interaction

It is helpful for those facilitating any group to develop certain skills in dealing with people and to be aware of and understand some fundamentals of basic human nature. Honing these skills can be useful in conducting efficient and productive meetings. As these skills are concerned with interpreting human behaviors, they are largely subjective in nature. And while none of these skills provide foolproof, objective, repeatable, or accurate results, they do provide a foundation for the facilitator to understand what motivates people, deal with certain behaviors exhibited by individuals, and guide a group to effective and positive outcomes. The following sections provide some techniques for reading people, including verbal and nonverbal communications, and for identifying types and ways to work with difficult people.

2.5.1. Reading People. Being able to "read people" involves paying careful attention to both the verbal and nonverbal communication taking place during any encounter between human beings. It is important to understand that being able to read people is not an exact science and research shows that most people who believe they read people well, typically do not. This could be in part due to their own weakness in understanding 18 and their actual capabilities and limitations, but typically the complexity of human behavior does not lend itself to accurate characterization. That said, there are some verbal and nonverbal behaviors that generally indicate certain moods, attitudes, and personality traits. Understanding and recognizing these will help the facilitation process for any group (or individual) interaction.

2.5.1.1. Verbal Communications. Effective communication is important to the success of any group interaction. The ability to exchange ideas, understand other perspectives, solve problems, and achieve goals depends significantly on how effectively we communicate with others.

Windle and Warren (n.d., accessed March 2014) discuss three components of effective communication: verbal, paraverbal, and nonverbal messages. This section focuses on the verbal and paraverbal aspects of communication and how each part, along with nonverbal signals, impacts our ability to effectively communicate. Nonverbal communication will be discussed in depth in the following section. The three components of communication are defined as follows:

- Verbal. Aspects of verbal messages are word choice, word arrangement, and message content.
- Paraverbal. Paraverbal messages deal with how words are stated. Aspects of paraverbal messages are tone, pitch, and speed.
- Nonverbal. Nonverbal messages are communicated through body language.

Effective verbal messages are clear, concise, and cogent. Listening to a rambling, unorganized speaker is tedious. Lengthy and convoluted dissertations not only lose the messages' relevance, but they also confuse listeners. Consider that Abraham Lincoln's Gettysburg address, which followed other presentations that day, is now regarded as one of the greatest speeches in American history. In just over two minutes, Lincoln reminded the audience of the principles decreed by the Declaration of Independence and campaigned for the continued preservation of the Union. Compare that to Edward Everett's two-hour, 13,607-word oration, presented prior to Lincoln's address, now seldom read, and only remembered for its length. The point is to choose words carefully, avoid slang and jargon, minimize acronym use, and eliminate superfluous information from the message. Additionally, refrain from using words that are critical, judgmental, sarcastic, or accusatory as it tends only to instill defensiveness in the person they are directed to. Defensiveness is not conducive to problem solving and achieving the goals of the group interaction.

According to Windle and Warren (n.d., accessed March 2014), paraverbal messages account for about 38% of what is perceived and understood by others. Consider the saying, "It's not what you say, it's how you say it." When the emphasis is placed on different words in the same sentence, the meaning of the sentence changes. For example:

- "**I** didn't say he was responsible." (It wasn't me)
- "I didn't **say** he was responsible." (I conveyed it some other way)
- "I didn't say he was **responsible.**" (I said something else)

As noted above, there are three major components of paraverbal messages (pitch, tone, and speed of words). Pitch is simply defined as the key of one's voice. A high pitch is often interpreted as anxious or upset. A low pitch sounds more serious and authoritative. This was so important to UK Prime Minister Margaret Thatcher that she worked with a vocal coach to lower her naturally high-pitched voice. Because people pick up and respond to it, practice varying voice pitch to add emphasis to various aspects of the message and keep the audience interested. No one enjoys listening to a monotone speaker.

Tone is the second component of paraverbal messages. Tone is produced through a combination of pitches which create a mood. Create a positive, authoritative tone by lowering pitch, smiling, sitting (or standing) straight and actively listening, and by controlling inner thoughts. Negative thoughts are reflected in the tone of voice.

The third component, speed, also effects communication. Someone speaking quickly is harder to understand than someone speaking at a moderate pace. On the other hand, speaking very slowly may result in a loss of interest on the part of the audience. Combine this with a monotone pitch, and the message (and perhaps the speaker's credibility) may be completely lost. Speed also affects the tone and quality of the message. A fast pace makes the communication rushed. Slow paced messages may be perceived by the listeners as unimportant. A moderate pace is the easiest for the listeners to focus on.

Given that communication is comprised of verbal, paraverbal, and nonverbal messages, it is important that these three messages be consistent. In cases of conflicting messages, it is the paraverbal and nonverbal messages that are most often believed.

A good rule of thumb as articulated by Dr. Robert Sutton states, "Speak as if we are absolutely correct, and listen as if we are absolutely wrong."

2.5.1.2. Nonverbal Communication. Research shows nonverbal behaviors make up a large percentage of daily interpersonal communication. Windle and Warren (n.d., accessed March 2014) claim that as much as 55% of what is perceived by others is through nonverbal communication. Given that so much

information is communicated nonverbally, it is useful for the facilitator to recognize nonverbal signals and understand what they potentially mean. Remember, interpreting nonverbal communications, like any other method of "reading people," is not an exact science and can be subject to misinterpretation.

Nonverbal communication takes its form mainly in two areas of a person: the face and the body. Emotions and moods are expressed through both facial expressions and body language. Cherry describes eight major nonverbal behaviors and the emotions expressed by these behaviors:

- Facial expressions: Facial expressions are responsible for a significant portion of nonverbal communication and convey several emotions. Universal facial expressions are those that are similar throughout the world and communicate the same emotion. These include happiness, sadness, fear, and anger. Other facial expressions/emotions include surprise, disgust, confusion, excitement, desire, and contempt.
- Gestures: Gestures are deliberate movements and signals used to communicate information. Some of the most common gestures include waving, pointing, and using one's fingers to indicate numeric amounts. Other gestures are arbitrary and often relate to specific cultures.
- Paralinguistics: Paralinguistics refers to vocal features that accompany speech and contribute to communication, but are not generally considered to be part of the language system. Examples of vocal features include vocal quality, loudness, inflection, pitch, tempo, and tone of voice. It may also include facial expressions and gestures.
- Body language and posture: While posture and movement convey information and indicate feelings and attitudes, research suggests body language is more subtle and less definitive than popular belief. Unfortunately, the media has focused on over interpretation of defensive postures such as arm-crossing and leg-crossing.
- Proxemics: Proxemics deal with personal space requirements and the role it plays in communication and social interaction. How far apart individuals stand during a conversation generally depends on the degree of intimacy between them. Other factors that influence the amount of space one needs or perceives to possess include social norms, the specific situation, and personality characteristics. Most people are familiar with the "close talker."
- Eye-Gaze, blinking, pupil size: While a person looking directly into another's eyes during a conversation indicates interest and attentiveness, prolonged eye contact can feel threatening. On the other hand, breaking eye contact may indicate distraction, uncomfortableness, or the concealment of true feelings. Generally, the rate of blinking increases and pupils dilate when people encounter things they like. As poker players do, a person may strive to conceal his or her feelings by trying to control eye movement.
- Haptics: In medicine, haptics refers to the science that deals with the sense of touch. In terms of non-verbal signals, haptics refers to communication through touch. Common emotions communicated through touch include affection, familiarity, and sympathy.
- Appearance: Finally, Cherry considers the choice of color, clothing, hairstyles, and other factors regarding appearance as nonverbal communication behaviors. Appearance can impact physiological reactions, judgments, interpretations, and first impressions. Additionally, different colors evoke different moods. Therefore, it is important for the facilitator to keep his or her own appearance in mind when interacting with teams in addition to understanding how his or her own perceptions may be influenced by the appearance of others.

Cherry also provides her list of the top ten nonverbal communication tips in her paper titled "*Master the Art of Nonverbal Communication with these Tips.*" Paraphrased below, Cherry describes each of these nonverbal communication behaviors and provides insight into their potential meaning:

- Pay attention to nonverbal signals such as eye contact, gestures, posture, body movements, and tone of voice. All of these can transmit important information not put into words.
- Look for incongruent behaviors. Listen for words that fail to match a person's nonverbal signals. For instance, an individual may be frowning at the same time he or she is claiming to be happy. Incongruent behaviors tend to mean the meeting message is being ignored. The individual's focus is likely on unspoken moods, thoughts, or emotions.
- Concentrate on tone of voice. Pay attention to how tone affects others. Use tone of voice to emphasize ideas or thoughts one wants to communicate.
- Use good eye contact. A person who does not make eye contact may be evading or hiding something. However, too much eye contact may appear confrontational and intimidating. Do not stare intently into someone's eyes. Intervals of eye contact lasting only four or five seconds is recommended by some communications experts. The facilitator should apply sound judgment regarding the appropriate amount of eye contact for each situation.
- Ask questions about nonverbal signals. Repeat back your interpretation of what was said and ask for clarification. Clarification should be asked for in a genuinely inquisitive and forthright manner. Avoid cynical or aggressive tones that might imply the individual is somehow "wrong." Examples of asking questions for clarification include:
    - "So, what you're saying is that …."
    - "Let me make sure I understand what you said…."
- Use signals to make communication more effective and meaningful. Both verbal and non-verbal communication work best to convey the message. Use body language that reinforces the message. This is particularly useful when making presentations or speaking to large groups.
- Look at signals as a group. A single gesture may mean many things or nothing at all so do not place too much emphasis on just one signal. Look for groups of signals that reinforce a common point.
- Consider context. Always consider the situation and context in which any communication occurs. Are the nonverbal communications appropriate for the context? More formal behaviors required in some situations might be interpreted differently than the same behavior performed in other settings. Concentrate on making signals match the level of formality.
- Be aware that signals can be misread. The firm versus weak handshake – neither may mean what you think. Always look for groups of behavior. A person's overall demeanor communicates more information than a single gesture.
- Practice, practice, practice. Practice these tips to build communication skills and the ability to correctly interpret signals from others. Always pay careful attention to nonverbal behavior.

2.5.2. Recognizing and Working with Difficult People. Working with difficult people is always a challenge. Difficult people come in a variety of forms including those that are perpetually negative and pessimistic, those that are toxic or hostile, those that are neurotic and anxious, and those with overinflated egos. Certain qualities such as meanness and a sense of worthlessness make some people consistently hard to handle. Additionally, some people have hair-trigger defensiveness that degrades their ability to listen and communicate effectively. These qualities lead people to bulk up self-esteem by putting down others. Most

likely, everyone has had to deal with difficult individuals, including leaders at some point in their lives. Most people are familiar with the bullies, the abusive, the self-serving, the arrogant, the screamers. Marano (2012) describes four specific types of difficult people and provides some suggestions for dealing with them:

2.5.2.1. The Hostile. Some common traits of the hostile person include being disagreeable, cynical, and mistrustful. Additionally, the hostile person has a highly explosive reactivity when confronted and always hates to be wrong. Hostile types include the "bully boss" and the passive aggressive individual whose typical modus operandi is to "throw people under the bus." When it is necessary to confront a bully directly, remain calm and professional. Never confront a bully in public as bullies will never back down in front of an audience. Tell the bully that his or her behavior is unacceptable and be specific about exactly what behaviors are at issue. Clearly explain to the bully how to treat others. There is no guarantee that these tips will get any positive results, but engaging in the same type of hostile behavior as that of the bully is guaranteed to get negative ones.

2.5.2.2. The Rejection Sensitive. The rejection-sensitive person deems all slights intentional and constantly scans for them, both real and imagined. They are unnerved in the face of any slight no matter how small. In very extreme cases, their behavior may include stalking (typically male). Rejection sensitive people have very little self-esteem and a lot of self-doubt. Rejection or the expectation of it makes these individuals hostile, although this aggression is generally passive rather than overt. Although challenging, remaining calm and keeping one's own reactivity low is the best way to deal with a rejection-sensitive person. Listen well to understand the individual and respond clearly to avoid the conversation from spiraling out of control.

2.5.2.3. The Neurotic. The neurotic person is usually a pessimist and often suffers from anxiety. Obstructionism is a common trait among neurotic individuals. They are cynical and tend to delay progress while dismissing the ideas of others. Dealing with the neurotic person requires maintaining a calm presence. Resist the temptation to write the difficult person off and try to understand his or her perspective without advocating it: "My experience has been different…."

2.5.2.4. The Egoist. Common traits of the egoist include the inability to compromise, insisting on being seen as "right," taking everything personally, and promoting his or her own interests first. For egoists, "It's my way or the highway." Egoists are inclined to respond strongly, even angrily when their desires are not met. The egoist may be the most difficult type of hostile, toxic person to deal with because of his or her narcissism and inability to compromise. Like the rejection-sensitive, the egoist is handled by remaining calm and keeping one's own reactivity low.

2.5.3. Additional Tips for Working with Difficult People.

2.5.3.1. Defusing a Difficult Encounter. In addition to the tips for dealing with specific types of difficult people identified above, Marano (2012) provides several tips recommended by physician and *Psychology Today* blogger, Susan Biali, to defuse a difficult encounter. First, minimize time with problem people by keeping interactions as short as possible. When interacting with a toxic person, keep the discussion or disagreement logical. Provide fact-based communication with minimal details only. Maintain focus on the hostile person during the conversation to help avoid being the target of demeaning comments, twisted words, or manipulation. If possible, avoid topics that may invite trouble. Accept the person as is; he or she will never be the person one would like him or her to be. As much as possible, refrain from trying to explain oneself as the hostile person will not empathize with others or see their point of view. Conducting an interaction with a hostile person around some recreational activity or entertainment may also help to soften or neutralize a problematic encounter.

2.5.3.2. Negative People. Kruse describes eight techniques to deal with individuals he likens to the *Saturday Night Live* character, Debbie Downer. Kruse postulates that some people, like Debbie, are only happy when they are unhappy and bringing down everyone else around them. These people are surrounded by negative energy that tends to infiltrate the moods of others. To prevent this, Kruse provides the following advice:

- Do not get dragged down. Do not let the "Debbie Downers" pull oneself into their world of negativity. The negative "vibes" emanating from the "Debbie Downers" of the world are not healthy and can hinder productivity. Misery may love company but avoid becoming the companion. Stay positive and focused on the objectives.
- Listen. While tempting, do not tune negative people out. Although their very nature is generally negative, there may be some solid thoughts or ideas in their blustering that can be useful to the group. Use good, normal listening techniques to extract those nuggets and change the attitude from a negative tone to a positive one.
- Use a time limit for venting. The occasional need to vent does not equal a perpetual pessimist. If individual(s) in the group need to vent, allow only 5 minutes or so and then move forward with the agenda by saying something to the effect of, "I understand your concerns and/or issue but we need to move on now. Perhaps we can address this again at some later date."
- Do not agree. Do not appease the "Debbie Downer" just to make him or her shut up and go away. Agreement only encourages the complaining.
- Do not stay silent. Staying silent even though actively listening will lead the difficult person to interpret one's silence as agreement. Others, if present, might assume this as well.
- Switch extremes into facts. Negative people often speak in extremes by frequently using terms such as "never" and "always." Most often perceived issues, problems, or slights should not be in terms of "never" and "always." Switch the negative person to fact-based statements only.
- Move to problem solving. Complainers frequently feel powerless and that most situations are hopeless. The best way to deal with this is to try to move them from continual complaining into problem solving.
- Cut them off. If all else fails, and after enough venting has been allowed to take place, one may just have to politely shut the difficult person down and move on to something else.

2.5.4. An Approach for Dealing with a Difficult Person in the DWT. The information above will help the facilitator recognize and deal with one or more difficult DWT members. If a difficult individual is negatively impacting the progress of the meeting, try the following steps (these should be discussed and agreed upon by the team lead and facilitator prior to the document writing event):

- Step 1. Discuss the situation with the DWT lead. If the team lead agrees that there is an issue and action should be taken, the facilitator should take the person aside during a break and speak frankly with him or her to resolve the issue. This may require one or more consultations between the facilitator and the disruptive individual. Always attempt to solve at the lowest level first. In most cases, the person needs to feel "heard" and the facilitator can do that without it disrupting the rest of the group. It is more effective for the facilitator to take this role because the group will view the facilitator as a neutral and more objective team member. The DWT lead is a stakeholder with his/her own interests and could potentially be perceived as less objective.
- Step 2. If Step 1 does not work, the facilitator should recommend to the DWT lead that he/she dismiss the individual and replace him/her with a representative from the same organization.

Keep in mind that there are some things that are simply out of the facilitator's control. Not all people are easy or pleasant to deal with. Recognize when the fight is not worth it and back off. Understand the things that can be changed and do not waste time and effort trying to change those that cannot be changed. At times it can be very difficult but strive to maintain professionalism and civility in these challenging settings. Know what a critical issue is and what is not. Finally, remember the maxim, "Perfection is the enemy of good enough; but never accept a "compromise" in a critical area just to end a contentious discussion."

SECTION 3. PREPARING FOR THE DOCUMENT WRITING EVENT

Since preparation is essential to the success of the DWT, this chapter describes five key elements to consider when preparing for a document writing event: initial communication with the DWT lead, developing the DWT objectives, determining DWT membership, developing the agenda and schedule, and identifying administrative and logistical considerations.

3.1. Initial Communication with the DWT Lead. The initial communication with the DWT lead is very important since it forms his or her initial impression of you as a facilitator, helps establish rapport, and enables the facilitator to determine the level of readiness to conduct the document writing event. This is a critical step in the process since the facilitator will be working closely with the team lead throughout the document writing event.

In preparing for the initial conversation, the facilitator will need to coordinate a date and time (most likely, several times) with the DWT lead to discuss the upcoming document writing event. This initial conversation may be an hour or more in length, so the facilitator and DWT lead should plan accordingly. The facilitator should prepare a list of questions beforehand to gain insights into various aspects of the DWT such as the team members, experience levels, participating stakeholders, tasks accomplished, and projected timeline. Do not assume all this has been done—be prepared with some thoughts about how to fill in any blank areas. Having a successful first encounter requires both people to have done their homework.

Although not an all-inclusive list of questions, Table 3-1 lists questions the facilitator should consider for his or her initial conversation. Based on the responses received from the HPT lead, the facilitator can assess DWT readiness, determine what additional actions must be taken to prepare for the document writing event, and begin formulating an approach to facilitate the DWT.

Table 3.1. Examples of Initial Conversation Questions

Topics	Questions
Experience, Background	What is your experience with conducting or participating in requirements development? What is your background (AFSC, past assignments, accomplishments)? What is your current job title and what responsibilities do you have?
Mission Area Knowledge	What is known about the mission area under study? What background documentation can you send to me so I can become familiar with subject? What documents exist today? Was there analysis that directly led to this? Can I get copies of them?
OAS Advisor	Contact OAS for assistance or advice in conducting the DWT as needed.
DWT Familiarity	How familiar are you with a Document Writing Team? Have you ever served as a leader or member of an DWT? If so, what DWT(s)?
Guidance/Process Knowledge	How familiar are you with the JCIDS manual, Capability Development Guidebooks, and other guidance relevant to your solution pathway? Do you have any questions regarding capability development, the JCIDS process or conducting a DWT?
SPR Products	When was the SPR conducted? What are the important aspects of the strategy described in the SPR? How does the planned DWT effort align with the SPR?

Air Staff Engagement	Have you spoken with the CDT functional representative at AF/A5/7? If so, who? Have you talked to anyone else in Air Staff? If so, who? What do they want from the study? Have they articulated any issues, key questions, scope, or other study requirements?
MAJCOM Engagement	What other directorates/divisions/offices in your MAJCOM have you collaborated with regarding this effort? Have they articulated any issues or key questions.
Document Writing Event Goals	What are the goals of the document writing event (i.e., develop draft document, develop final document)? How many days do you think is needed for the document writing event? What administrative support do you have? What assistance do you need in planning and arranging the document writing event?
DWT Members	What organizations should have DWT membership? How many members do you think you need for the DWT? Who have you already invited for DWT membership? Who are you considering for DWT membership? What experience do the selected members and those you are considering have in developing requirements? What expertise do they have? What expertise is needed? What assistance do you need in forming the DWT?
TDY Funding	Which DWT participants need to travel? Do they have funding? Have you considered virtual participation?

3.2. Solution Pathway Review. As described in AF/A5/7 Capability Development Guidebook series, a Solution Pathway Review (SPR) is required for all Air Force-sponsored programs entering the requirements process regardless of where the program enters the process. The SPR is conducted by the AF/A5D before the sponsor convenes the DWT. The SPR provides a cross functional, corporate evaluation of identified requirement gap(s) and determines the best solution pathway to address the identified gap(s). Before talking with the DWT lead, the facilitator should review the SPR sections in the applicable Capability Development Guidebook(s). In addition, the facilitator should advise the DWT lead and members to review the applicable guidance before the document writing event.

In preparing for the document writing event, the facilitator should ensure the DWT lead understands the information from the SPR. The facilitator should advise the team lead to maintain alignment with the SPR in developing the requirements document.

3.3. Key Planning Factors to Consider. A major challenge for the facilitator and DWT lead is determining the length of the document writing event and the tasks that will be accomplished on each day of the event. There are several key planning factors the facilitator, in collaboration with the team lead, must consider when preparing for the document writing event. At a minimum, the facilitator must assess the level of experience of the team, the complexity of the problem, and the amount of work that has been accomplished when determining the length of the document writing event and the tasks to be accomplished on each day. For instance, a more experienced team that has developed a good quality initial draft of the requirements document on a less complex problem will require less time to complete tasks, so the length of the event will likely be short. In contrast, a less experienced team that has developed a very rough and largely incomplete draft of the requirements document on a complex problem will require more time to complete the document writing event tasks, so the length of the event will likely be longer.

In most cases, tasks tend to take longer than planned so the facilitator should keep this in mind when allocating time to accomplish the document writing tasks.

3.4. Develop DWT Objectives. A well-thought-out set of objectives is essential to the success of the DWT. The facilitator should work with the team lead to ensure that the objectives are established, documented, realistic, and clearly articulated to DWT members prior to beginning the document writing event.

The following should be considered as early in the process as possible:

- Determine length of the event. As noted in the previous section, the facilitator, in collaboration with the DWT lead, considers several key planning factors to determine the appropriate length of the event and tasks to be accomplished on each day.
- Help the team lead determine the list of attendees (organization and/or individuals) and their function on the team. Review the capability development guidebooks and the SPR results to help determine the attendees. Funding may be a factor when determining how many people can attend in person or virtually.
- The facilitator will collaborate with the DWT lead to develop an approach for conducting the document writing event (e.g., length, breakout sessions, working group configuration, homework required prior to the event). The approach will depend upon many factors, including the nature of the problem, the amount of work accomplished already, and the anticipated members.
- Decide on the level of completion expected. Will the DWT deliver a draft or a near-final document? There should be shared understanding of the level of completeness expected at the end of the document writing event. If the team's homework has been properly accomplished and the right people are on the team (and empowered to speak for their organization in most things), then the goal should typically be a near-final document ready for coordination.
- Consider the administrative details early and make decisions about location, facilities, refreshments, accommodations, transportation, lunch options (working lunch, on your own). See Table 3.2. for a list of administration tasks to consider.
- Four suggestions for consideration:
    - Include an Icebreaker: An icebreaker activity is very helpful in allowing DWT members to meet and understand what each person brings to the table, their backgrounds and expertise, and who they need to get to know better during breaks. Much of the DWT work is done during breaks out in the hallway or over lunch.
    - Avoid an Overly Aggressive Schedule: Section 2.1 of this guidebook discusses the four stages of team evolution. The change from Stage 1 to Stage 2 typically requires time for people to think about it and to internalize it. The same is true from Stage 2 to Stage 3. Add in the fact that DWTs should be comprised of experts in their various areas, and such people do not quickly acknowledge that their ingoing opinions need to change—this too takes time for thought, reflection, and discussion. Groups of this type cannot produce quality products in artificially compressed timelines. They almost always take longer than the team lead desires.
    - Take Real Lunch Breaks: DWTs typically deal with problems that are complex and difficult to solve. Consequently, there will likely be some very contentious moments during the document writing event. This is to be expected and is a good thing. Taking real lunch breaks will allow members to relax, recharge, and take care of other business (e.g., calling back to home station, and working sidebar issues). The time spent in taking real lunch breaks will invariably be worth the time lost by not having working lunches. The best facilitators often try

to schedule the contentious pieces in the mid to late morning and therefore use the lunch break as a planned "cooling off" period.

- Impasse? - Consider a Working Dinner: This seems to contradict the point directly above, but it does not. It is not uncommon for a DWT to encounter a roadblock. For example, two members from two different organizations who are essentially stating and restating each organization's position over and over with neither side being able to back down in public and lose face. Without some shifting of these positions, progress will be stalled. Often a carefully planned dinner meeting is a way to break the impasse. The two members, along with two or three others they both trust and value, go to dinner. It is a less formal situation, but more importantly it provides a venue where some give and take is possible without either side publicly backing down. Often, the next morning, the team lead or facilitator can simply announce that a compromise has been reached and move on.

3.5. DWT Membership.

Determining DWT membership requires significant thought and deliberation. The SPR worksheet and prior efforts in the mission area provide good insight on key stakeholders and participants who should be included on the DWT. The objective is to assemble a team with varied cross-functional expertise. Consider adding representatives via an "A staff" structure example – A2 (Intelligence), A3 (Operations), A4 (Sustainment), A5 (Requirements), A6 (Communications). Additionally, technical expertise such as engineering support should be considered. Include additional functional area experts as needed such as Operational Test, energy, etc. Team composition normally includes a mix of both government and contractor personnel.

3.6. Developing the Agenda

The DWT lead collaborates with the facilitator to plan the agenda and communicate the information to the DWT members. An agenda is the framework that helps DWT run effectively and efficiently. It is a step-by-step outline of the topics to be covered at the document writing event. Effective agendas enhance group accomplishments:

- The agenda informs team members of accomplishments and priorities.
- It ensures adequate consideration of all issues, events, and projects.
- It identifies the order in which topics will be addressed.
- It keeps the discussion focused and on track.
- It focuses and encourages better pre-document writing event preparation.
- It makes effective use of participants' time.

A detailed agenda will help the facilitator communicate what needs to be accomplished. The facilitator should be very well prepared and well versed on the DWT process and mission area of interest prior to the meeting.

It is important to balance the time allotted with the agenda goals and objectives. For example, some tasks may be more difficult than originally thought. In these cases, the agenda may need to be revised if the team falls behind. Agendas can be flexible, but every effort should be made to accomplish the objectives set for the document writing event. If there are some objectives that must be accomplished during the event, the agenda should be structured to ensure the "must do" objectives will be accomplished. Some items may need be accomplished after the event. Establishing action items with OPRs, and suspense dates is a common approach.

3.7. Administration

There are many important details involved in planning and executing a document writing event that must be accomplished. Being organized and having a list of administrative considerations is critical when preparing to execute an event. The host, often the DWT lead, will likely delegate many of the administrative tasks to one or more individuals responsible for handling these tasks. Some administrative considerations are in Table 3-2.

Table 3-2. Administrative Considerations

Task	Subtasks Involved	Caveats
Obtain funding for DWT	Includes travel, refreshments, facility fees if applicable	Funding may be limited due to fiscal environment
Enlist DWT Participants	Contact stakeholder organizations, send invitation letter, obtain supervisor approval, ensure participants have relevant knowledge, skills, experience	Personnel limited. Start early
Publish team contact information	Collect, populate SharePoint site	Will be a living document
Determine Location	Obtain attendee list first to ensure most convenient and cost-effective location is chosen	Funding, meeting length, travel approval timelines, competing events in the area, local amenities
Obtain Required Technology	Ensure computers are compatible at correct security levels, projectors, copy machine, etc.	Potential delays in obtaining required resources
Security	Obtain SMO codes and passing instructions	Establish location first, if held at too high a level, some may be unable to participate
Designate Roles	Meet with core team to determine roles	Ensure roles are clearly defined
Create document library	Establish a shared location, populate, communicate location	Consider document classification, system availability
Generate Agenda	See section 3.6	Inadequate agenda results in decreased value of HPT
Arrange appropriate security precautions	Determine meeting classification level, arrange as appropriate	Side meetings are possible at higher levels with limited attendance
Meeting nourishment	Procure snacks, beverages, lunches	Decide if working lunch is necessary
Meeting equipment	Secure audio visual tools	Type of meeting

The success of the DWT will rely on several behind-the-scenes activities. This section is dedicated to introducing some of the activities that contribute to extracting the most from the highly talented group that may only have a short time allocated to contribute. Table 3.2 provides a list of critical tasks that must be accomplished to plan and conduct a document writing event.

There are numerous documents that provide regulation, policy, instruction, background, and technical information on the mission area and processes (Table 3-3). DWT members should review these documents in preparation for the document writing event. OAS recommends developing a collection of documents in a widely accessible central location such as a SharePoint site. Physical distribution of the material through email is more cumbersome but can accomplish the goal of the team understanding the governance, previous study work, and state of the art. Providing this information ahead of time will help prepare members for the event and enable them to be productive at the start. Regardless of the storage or distribution mechanisms, it is crucial that someone be personally responsible for configuration control of all the key documents the team will use and create. Whatever mechanism is chosen, be aware that classification issues and team members outside the Air Force may require special handling to make information available to all who need it.

Table-3.3: Important DWT Documents

What*	Why
Capability Development Guidebooks	Guidance on AF requirements development including roles, responsibilities, method, and deliverables
JCIDS Manual	JCIDS document usage, format, content, and staffing. AF guidance layers on top.
SPRs, predecessor requirement documents, and studies	Necessary to learn about mission area under study – prevent “reinventing the wheel”
Mission area background	Mission area knowledge is necessary to effectively facilitate and conduct a document writing event
Minutes and action items from SPR and pre-/post- document writing event meetings	Describes the key decisions and happenings of the document writing event and records the post-HPT way ahead and actions to be completed
*Documents should be made available to team members as soon as the library is established (see Table 3.2.)	

SECTION 4. CONDUCTING THE DOCUMENT WRITING EVENT

The document writing event agenda normally focuses on the review of a pre-meeting draft document (highly recommended) and often requires two or three days. Use of MS Word and MS PowerPoint on-screen enables discussion. Meeting location, attendees, security level and clearances, and agenda are worked-out in advance. Consider break-out session(s) at the end to work action items as needed.

Example Agenda: Tailor to situation as required:

- Facilitator: Establish team rules as in Table 4-1.
- Facilitator: Team introductions (background, specialty, role).
- Sponsor: Overview of the task at hand (mission area, agenda, objective).
- Facilitator: Line by line walk-though of each section of the draft document.
    - Team member brainstorming and real time edits to draft document.
    - Identify holes and establish action items as needed.
    - Build-in 15-minute breaks (every 2 hours), lunch breaks (1-1.5 hour).
- Sponsor: Closing remarks, recap (action items), plan for next day.

Table 4.1. Team Rules

Rule	Description
Active Participation	The team benefits when all members are contributing to the effort. Individual team members must actively participate and not rely on one or a few team members to accomplish the preponderance of work. There are many ways members can participate such as sharing the tasks, voicing positions or opinions (rather than remaining silent) and offering solutions to problems as they arise. Participation will not be uniform. The logistician will obviously be most engaged in the logistics related parts of the meeting but should stay engaged throughout.
Withhold Criticism	Withholding criticism, especially during brainstorming sessions, is necessary for encouraging creative thinking. Withholding criticism does not indicate support or agreement, but instead enables team members to generate ideas without fear of disapproval. For this to be effective everyone needs to understand “silence is NOT consent”, and time must be allotted to critically discuss any brainstorming list. A common problem is an unrealistic schedule that does not allow this critical discussion and that can result in brainstorming ideas being accepted into the final product not because they were good, but because time ran out.
Avoid Attribution	The documents produced are accomplished through a team effort. Attributing specific text or sections of text to a single team member or smaller group within the HPT can provoke criticism or jealousy and does not help the HPT achieve consensus. Sometimes the collective team wants to attribute something to a certain member because it increases the credibility based on that member’s credentials, but this should be the exception.
Strike a Balance	Perfection is the enemy of good enough, but so is being driven by schedule constraints and prematurely declaring “good enough”.

References

Cherry, Kendra, *Types of Nonverbal Communication; 8 Major Nonverbal Behaviors*, About.com Psychology, (http://psychology.about.com/od/nonverbalcommunication/a/nonverbaltypes.htm)

Cherry, Kendra, *Top 10 Nonverbal Communication Tips; Master the Art of Nonverbal Communication with these Tips*, About.com Psychology, (http://psychology.about.com/od/nonverbalcommunication/tp/nonverbaltips.htm)

Harkins, Phil, 10 Leadership Techniques for Building High Performance Teams (http://www.linkageinc.com) Copyright © 2006-2008 Linkage. All rights reserved.

Hallett, Tom. Mind Tools (http://www.mindtools.com/CommSkll/ActiveListening.htm) Mind Tools Ltd, (n.d., accessed March 2014).

Hersey, P. and Blanchard, K. H. *Management of Organizational Behavior—Utilizing Human Resources*. New Jersey. Prentice Hall. 3rd Edition, 1977.

Jackson, Bruce Jackson and Madsen, Susan R. "Common Factors of High Performance Teams" *Journal of Contemporary Issues in Business and Government* 11.2 (2005): 35-49.Available at: http://works.bepress.com/susan_madsen/84

Kruse, Kevin, *8 Tips for Dealing with Difficult People*, Forbes, (http://www.forbes.com/sites/kevinkruse/2013/06/25/dealing-with-difficult-people/)

Marano, Hara Estroff, *The High Art of Handling Problem People*, Psychology Today, May 01, 2012, (http://www/psychologytoday.com/articles/201204/the-high-art-handling-problem-people)

Tuckman, Bruce (1965). "Developmental Sequence in Small Groups", *Psychological Bulletin* 63(6): 384-399. Reprinted with permission in Group Facilitation, Spring 2001.

Windle, Rod and Warren, Suzanne (n.d)*, Communication Skills*, (http://www.directionservice.org/cadre/section4.cfm)

A Short Guide to Consensus Building, An Alternative to Robert's Rules of Order for Groups, Organizations and Ad Hoc Assemblies that Want to Operate by Consensus (n.d.) (http://web.mit.edu/publicdisputes/practice/cbh_ch1.html)

Appendix A - Measures

A Practical Guide for Developing and Analyzing Measures in the Capabilities-Based Assessment, pre-Materiel Development Decision Analysis, and Analysis of Alternatives

31 Jul 23

TABLE OF CONTENTS

1.0 INTRODUCTION

This chapter provides an introduction to measures by discussing the importance of measures, how measures are defined, and types of measures.

1.1. Importance of Measures

The Capabilities-Based Assessment (CBA), pre-Materiel Development Decision (MDD) analysis, and Analysis of Alternatives (AoA) require measuring various aspects of interest as part of the analysis. Measures are vital to the analysis since they provide the basis for the assessment and the conclusions drawn from the assessment. When properly developed and explicitly stated, measures will:

- Specify what to measure
- Determine the type of data to collect
- Identify how data is collected
- Identify resources required to perform data collection
- Identify how the data can be analyzed and interpreted

The CBA, as the analytic basis for capabilities requirements development, requires the development and analysis of measures to determine whether gaps exist in the baseline capabilities of the force. From these capability gaps, capability requirements are developed and potential solutions are identified to close or mitigate the gaps.

In the pre-MDD analysis, measure development and analysis varies based on the focus of the study. For example, a pre-MDD analysis with the purpose of scoping down the number of potential concepts in preparation for an upcoming AoA will require measure data collection and analysis to identify the most technically feasible concepts.

Note on pre-MDD Analysis

The CBA can be categorized as a pre-MDD analysis since it is conducted before the materiel development decision (MDD). The term "pre-MDD analysis" in this handbook refers to a study that is accomplished after the CBA and prior to the MDD. The purpose of the pre-MDD analysis can vary, but typically it is used to further refine the requirements strategy for the capability gaps identified in the CBA or to shape and scope the AoA.

In the AoA, measures are developed and used to assess alternatives and their potential to meet capability requirements. Measures are essential for comparing the performance of alternatives, determining how well they close or mitigate capability gaps, and identifying the best-value alternative through cost-capability analysis.

Measures that are developed and assessed in the CBA, pre-MDD analysis, and AoA serve as the analytic foundation for developing capability requirements. Measures in a CBA and pre-MDD analysis conducted before the Initial Capabilities Document (ICD), for example, can be used to develop capability requirements in an ICD and to determine whether one or more Joint or Air Force DOTmLPF-P Change Requests (DCRs) should be initiated. Furthermore, measures in all of these studies can be used to develop capability requirements in a Capability Development Document (CDD) and CDD Update (formerly CDD UPDATE). The measure analysis conducted in these studies underpins the development of Joint

Capabilities Integration and Development System (JCIDS) and Air Force capability requirements documents.

Good measures enable the analyst and study team to accurately and concisely interpret and report results of the analysis. Measures that are ambiguous or do not measure the right attributes of interest in the study make it difficult, if not impossible, to interpret and draw conclusions from the results of the analysis. Poorly conceived measures can be detrimental to meeting the study objectives and negatively impact the credibility of the analyst, study team, and study sponsor.

1.2. What is a Measure?

A measure is a device designed to convey information about an entity being addressed. It is the dimensions, capacity, or amount of an attribute of an entity of interest in the analysis. An attribute is a quantitative or qualitative characteristic of an element or its actions. Survivability, persistence, availability, and accuracy are examples of attributes.

Attributes of tasks form the basis for developing measures. A measure is used to describe varying levels of an attribute and to provide the foundation for comparison. Measures are not requirements, conditions, or criteria, but are developed from requirements, measured under conditions, and evaluated against criteria.

Note on Analysis

The term "analysis" is used throughout this handbook and is defined as the categorizing, ordering, manipulating, and summarizing of data to gain insights needed to answer study questions. Through analysis, the analyst organizes data into an intelligible and interpretable form to make inferences and draw conclusions.

1.3. Types of Measures

There are many different types of measures that have been developed for various purposes. Though various types of measures have been used in CBAs, pre-MDD analyses, and AoAs in the past, there are three types of measures that are commonly used:

- Measure of Effectiveness (MOE)
- Measure of Suitability (MOS)
- Measure of Performance (MOP)

A measure associated with an attribute of operational effectiveness is referred to as a measure of effectiveness (MOE). Operational effectiveness is the overall capability of a system to accomplish a mission when used by representative personnel in the environment planned or expected for operational employment of the system considering organization, doctrine, tactics, survivability, vulnerability, and threat. The MOE is defined as:

> *The data used to measure the military effect (mission accomplishment) that comes from using the system in its expected environment. That environment includes the system under test and all interrelated systems, that is, the planned or expected environment in terms of weapons, sensors, command and control, and platforms, as appropriate, needed to accomplish an end-to-end mission in combat.* (DAU Glossary)

A measure associated with an attribute of operational suitability is referred to as a measure of suitability (MOS). Operational suitability is the degree to which a system can be placed satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human systems integration, manpower supportability, logistics supportability, natural environmental effects and impacts, documentation, and training requirements (JCIDS Manual). The MOS is defined as:

> *Measure of an item's ability to be supported in its intended operational environment. MOSs typically relate to readiness or operational availability and, hence, reliability, maintainability, and the item's support structure.* (DAU Glossary)

A measure associated with system characteristics and performance is referred to as a measure of performance (MOP). The MOP is defined as:

> *System-particular performance parameters such as speed, payload, range, time-on-station, frequency, or other distinctly quantifiable performance features. Several MOPs may be related to achieving a particular Measure of Effectiveness (MOE).* (DAU Glossary)

2.0. UNDERSTANDING MEASURES AND DATA

2.1. Levels of Measurement

Table 2-1 shows the four general levels of measurement: nominal, ordinal, interval, and ratio. These range in sophistication from low (nominal) to high (ratio). Data associated with nominal and ordinal levels of measurement are commonly referred to as qualitative data. Data associated with interval and ratio levels of measurement are referred to as quantitative data. Since data characteristics are different at each level, there are particular statistics that are appropriate for each level.

The analyst must consider various factors such as the attribute being measured, purpose of the measurement (e.g., counting objects in categories, attaining a rank order), and data collection requirements when determining the levels of measurement that will be used. The analyst should strive to use the highest levels of measurement that are possible and suitable for the study. It is important that the analyst understand the levels of measurement to ensure the appropriate statistics are used.

Note on Analysts

The term "analyst" is used throughout this handbook and refers to individuals or team members assigned to conduct all or some of the analysis in a study. In some cases, different analysts with a range of experience and expertise may be involved in various parts of the analysis. The analyst is responsible for organizing data into an intelligible and interpretable form to make inferences and draw conclusions.

Table A-2-1: Levels of Measurement

Level	Description	Appropriate Statistics	
		Descriptive	Inferential
Nominal	Data are assigned the same symbol if they have the same value of the attribute. *Example: 1 – Male, 2 – Female*	Mode, percentages, frequencies	Chi-square, binomial, McNemar, and, Cochran Q tests
Ordinal	Data are assigned numbers/symbols such that the order of the numbers/symbols reflects an order relation based on the attribute. *Example: 1 – Good, 2 – Better, 3 – Best*	All statistics permitted for nominal scales plus percentile (e.g., median (50th percentile), 80th percentile, 95th percentile)	Mann-Whitney U- test, Kruskal Wallis test, Friedman two- way analysis of variance, rank-order correlation
Interval	Data are assigned numbers such that differences between numbers represent equivalent intervals. *Example: Temperature in degrees Fahrenheit*	Statistics permitted for ordinal plus mean, standard deviation, and range	Product-moment correlation, Z-test, T- test, F-test, factor analysis, ANOVA
Ratio	Data are assigned numbers that have the features of interval measurement and meaningful ratios between arbitrary pairs of numbers. There is a rational zero point for the scale which is necessary for the ratio statements to have meaning. *Example: Length in feet; duration in seconds*	All statistics permitted for interval scales plus geometric mean and harmonic mean	Same as interval plus coefficient of variation

Derived from: Kerlinger (1986); Leedy (1997); Tull and Hawkins (1980); Churchill (1979); Zikmund (1991)

2.2. Measure Description

As shown in Table 2-2, there are seven elements that are typically used to describe a measure. Measures are developed in the context of the tasks and attributes of interest in the study. A task, also referred to as a mission task, describes what is expected to be performed. As noted previously, an attribute is a quality or feature of something that is relevant to the task. Together, the task and attribute form the basis for developing the measure.

Using the example in Table 2-2, accuracy is an attribute of the strike target task. The measure statement describes miss distance as the measure that will be used to determine how accurate a system performs in striking a target.

Table 2-2: Example of a Measure Description

Task	Attribute	Measure	Metric	Criteria	Data
Strike Target	Accuracy	Miss distance	90th Percentile	Threshold: ≤ 5 meters Threshold = Objective	Distance from the intended point of impact to the actual point of impact
Conditions:	Time of day (night-time, day-time); Weather (instrument meteorological condition, visual meteorological conditions); Terrain (mountainous, plateau)				

The other elements of the measure description include the metric, criteria, data, and conditions. Table 2-2 shows examples of these elements. Each element is defined below:

Metric: a unit of measure that coincides with a specific method, procedure, or analysis. The mean, median, mode, percentage, and percentile are examples of a metric.

Criteria (also referred to as standards): define the acceptable levels or standards of performance for a metric and are often expressed as a minimum acceptable level of performance (threshold) and desired acceptable level of performance (objective).

Data: individual measurements that are used to compute the metric for a measure.

Conditions: describe the operational environment in which the task will be performed.

2.3. Measure Criteria

Measure criteria (standards) describe threshold and objective levels of performance that are based on capability requirements. The difference between the threshold and objective values sets the trade space for balancing multiple performance attributes and parameters. The threshold and objective are defined as follows:

Threshold: a minimum acceptable operational value of a system capability or characteristic below which the utility of the system becomes questionable.

Objective: an operationally significant increment above the threshold. An objective value may be the same as the threshold value when an operationally significant increment above the threshold is not identifiable.

The threshold and objective terms and associated definitions apply to both the measures developed in the CBA, pre-MDD analysis, and AoA and the capability requirements expressed in the CDD and CDD Update. Values used with measures developed in the CBA, pre-MDD, and AoA can serve as the basis for developing threshold and objective values for capability requirements in the CDD and CDD Update.

In the ICD, capability requirements do not specify a threshold or objective value, but instead specify an initial objective value. The intent of the initial objective value is to provide a starting point that not only satisfies an operational need, but also enables the analysis of capability requirement tradeoffs above and below the initial objective value. The measures and associated threshold and objective values developed in the CBA and pre-MDD analysis (those conducted before the ICD) can serve as the basis for developing the initial objective values of capability requirements in the ICD.

There are two types of measure criteria: user-established and identified. User-established criteria are criteria that are explicitly stated or implied in a capability requirements document (ICD, CDD, and CDD Update). When user-established criteria do not exist, criteria must be developed as part of the study to enable the analyst to assess the measure. Criteria that are developed are referred to as identified criteria. Section 3.4 (Identify Measure Criteria and Metrics) provides additional information about user-established and identified criteria.

2.4. Measures and High Interest Parameters and Attributes

High interest parameters and attributes known as Key Performance Parameters (KPPs) and Key System Attributes (KSAs) define capability requirements that are critical or essential. KPPs and KSAs are identified in the CDD and CDD Update. CBAs, pre-MDD analyses, and AoAs provide the analytic foundation for determining which parameters and attributes should be KPPs and KSAs. Measures developed in these analyses serve as the basis for identifying the most critical or essential aspects of capabilities and developing recommendations for potential KPPs and KSAs and their associated threshold and objective values. KPPs and KSAs are defined as follows:

> *Key Performance Parameter (KPP): performance attributes of a system considered critical or essential to the development of an effective military capability.* (JCIDS Manual)
>
> *Key System Attribute (KSA): performance attributes considered important to achieving a balanced solution/approach to a system, but not critical enough to be designated a KPP.* (JCIDS Manual)

For all capability solutions being developed, there are mandatory KPPs & KSAs that must be addressed by the study sponsor whether relevant to the capability or not (JCIDS Manual). It is important that the analyst understand the nature of the systems, concepts, or alternatives being analyzed in the study to determine which KPPs are relevant. In addition to the mandatory KPPs and KSAs, the study team may identify other development parameters and attributes that should be considered as KPPs and KSAs. By knowing the relevant mandatory KPPs and any additional development KPPs and KSAs identified by the study team, the analyst can construct appropriate measures, collect the right data, and conduct the analysis.

Performance attributes that are not important enough to be considered KPPs or KSAs, but still appropriate to include in the CDD or CDD Update are designated as Additional Performance Attributes (APAs). As is the case for KPPs and KSAs, measures developed in the CBA, pre-MDD analysis, and AoA can be used to identify potential APAs and their associated threshold and objective values.

Finally, other system attributes (OSAs) are used to identify any other attributes not previously identified, especially those that tend to be design, life cycle cost, or risk drivers. Some examples include physical or operational security needs, transportability, deployability, human systems integration considerations, and

space, weight, power, and cooling requirements (see the JCIDS manual for additional information). Measures developed in the CBA, pre-MDD analysis, and AoA can be used to identify potential OSAs and their associated threshold and objective values.

Note on Measures and Capability Requirements

Although related, measures and capability requirements serve different purposes. Measures developed in the CBA, pre-MDD analysis, and AoA serve as the analytic basis for developing capability requirements. A capability requirement is a capability that is required to meet an organization's roles, functions, and missions in current or future operations. Capability requirements are described in capability requirements documents (ICD, CDD, and CDD Update). An ICD specifies one or more capability requirements and associated capability gaps which represent unacceptable operational risk if left unmitigated. In a CDD and CDD Update, capability requirements are specified in terms of KPPs, KSAs, APAs, and OSAs to support development of one or more increments of a materiel capability solution. During test and evaluation, measures are derived from KPPs, KSAs, APAs, and OSAs to facilitate the testing and evaluation of the materiel capability solution.

2.5. Determining How to Use Data

The analyst must determine what data is important enough to be measure data and how all other data will be used in the study. In addition to computing metrics for measures, the analyst can use data for other purposes such as inputs to models. As shown in Figure 2-1, altitude is an element of all these studies, but how it is used in each study is very different.

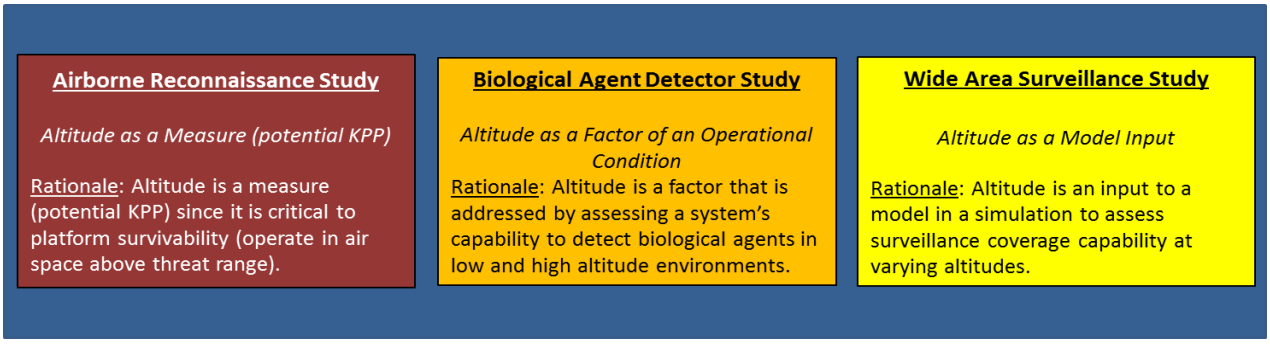


Figure 2-1: Examples of Using Data for Different Purposes

Although significant amounts of data may exist, the analyst must consider several factors when determining how to use data:

- Study guidance, objectives, questions, ground rules, assumptions, constraints
- Attributes of interest in the study
- Data collection requirements, availability of data, and confidence in data
- Capabilities of models or applications to produce measure values.

3.0. MEASURE DEVELOPMENT PROCESS

3.1. Process Overview

The measure development process consists of four major steps (Figure 3-1). The process may be iterative, meaning that as new information is learned it may be necessary to repeat previous steps. In addition, the process can be tailored by the analyst depending on his or her needs and the requirements of the study. The output of the process is a fully defined set of measures that links together the tasks, attributes, conditions, measures, criteria, metrics, and data that will be used in a study.

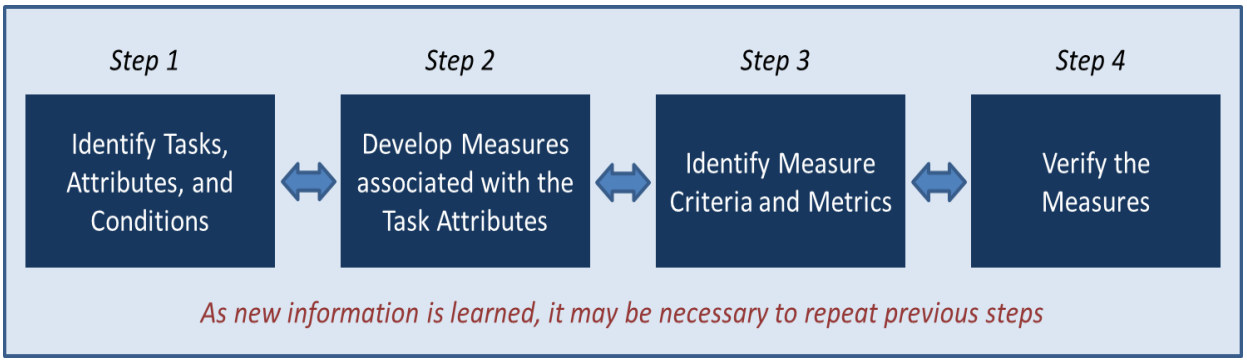


Figure 3-1: Measure Development Process Steps

3.2. Step 1: Identify Tasks, Attributes, and Conditions

The first step in the process entails identifying the tasks, attributes, and conditions. Tasks are derived from the mission that is expected to be accomplished. The mission is a statement of the action to be taken and the reason behind the action. Through mission analysis, the analyst, in collaboration with operational experts or subject matter experts, defines the requirement to perform tasks and the context of each task's performance to include the conditions under which a task must be performed. Mission analysis enables the analyst to gain an understanding of when and where a task must be performed and how the performance of a task contributes to mission success.

To conduct the mission analysis, the analyst should utilize the experience and expertise of subject matter experts knowledgeable of the operational concepts relevant to the mission area of interest in the study. Expert elicitation is a particularly useful method for deriving tasks from a mission and gaining insights into attributes, conditions, and measures that should be considered for each task (see Annex F for more discussion about expert elicitation). Although all experts will be knowledgeable of the mission area, they have different experiences and perspectives that will produce insights that may not be possible without their involvement.

Another important aspect the analyst must consider is the linkage to capability gaps and requirements. For the pre-MDD analysis and AoA, tasks and associated attributes and conditions should be linked to the capability gaps and requirements that are identified in the CBA(s) and capability documents such as the Initial Capabilities Document (ICD). The analyst must be able to show how the tasks and associated attributes and conditions are traceable to the capability gaps and requirements of interest in the study. Defining this linkage is the first step in determining how well capability gaps can be closed or mitigated, one of the main objectives of the AoA.

For the CBA, the analyst derives tasks, attributes, and conditions from capability requirements that are developed as part of the analysis. The analyst uses the capability requirements and the associated tasks, attributes, and conditions to determine whether capability gaps exist in the baseline capabilities of the force. The analyst must be able to show how the tasks and associated attributes and conditions are traceable to the capability requirements of interest in the study. For additional information on deriving tasks, attributes, and conditions for the CBA, see the AF/A5/7 Capability Development Guidebook, Volume 2C, Capability Based Assessment.

Table 3-1 provides an example that shows the linkage between the tasks, attributes, conditions, and the capability gap and requirement that would be appropriate for a pre-MDD analysis or AoA. As shown in the table, the capability gap describes a lack of global integrated intelligence, surveillance, and reconnaissance capability. An associated capability requirement is to provide moving target indicator support to maneuver and surface forces. From this requirement, the analyst derived three tasks with associated attributes and conditions in the context of the mission.

In addition to using expert elicitation, the analyst should conduct a literature review to gather information for identifying tasks, attributes, and conditions. Sources of information the analyst can use include the following:

- Joint Capability Areas (JCAs)
- Task lists (e.g., Universal Joint Task List (UJTL), Joint Mission-Essential Task List (JMETL), Mission-Essential Task List (METL), Air Force Task List (AFTL), other Service task lists)
- Support for Strategic Analysis (formerly known as the Analytic Agenda) documents (e.g., Defense Planning Scenarios (DPSs), Integrated Security Constructs (ISCs))
- Planning and operations-related documents (e.g., OPLANs, CONPLANs, CONOPS, CONEMPs)
- Concept documents (e.g., Concept Characterization and Technical Descriptions (CCTDs), Joint Concept Technology Demonstration (JCTD) reports)

When gathering information, the analyst should consider the following questions:

- What capability gap(s) are being addressed?
- What are the desired effects?
- What objectives, major operations, or activities are to be accomplished?
- What does the system do to support the mission?
- How will the system be employed?
- What are the key aspects of the operational environment the system will be employed in?
- What operational performance attributes (e.g., precision, responsiveness) and support attributes (e.g., compatibility, reliability) are described?

Table 3-1: Capability Gap Linkage Example

Capability Gap: Lack of global integrated intelligence, surveillance, and reconnaissance capability.		
Capability Requirement: Provide moving target indicator support to maneuver and surface forces.		
Tasks	Attributes	Conditions
Find target (detect, identify)	Accuracy, Timeliness	Time of day (night-time, day-time); Weather (instrument meteorological condition, visual meteorological conditions); Terrain (mountainous, plateau)
Track target	Accuracy, Persistence	Time of day (night-time, day-time); Weather (instrument meteorological condition, visual meteorological conditions); Terrain (mountainous, plateau)
Communicate information	Accuracy, Timeliness, Completeness	Electronic warfare environment (benign, contested)

It is common for one or more tasks to have dependent relationships with other tasks. In the Table 3-1 example, the "track target" task is dependent on the "find target" task. Without first finding the target, it is not possible to track the target. The analyst should understand these interdependencies and the potential capability tradeoffs that may warrant further analysis in the study.

Often there are multiple attributes that are associated with each task as illustrated in Table 3-1. When identifying attributes, the analyst should consider the most critical qualities or features that are relevant to the task. It is important to note that the number of attributes identified for each task can drive the scope of the study since each attribute will require at least one, and perhaps several, measures.

Annex D provides a list of attributes by Joint Capability Area (JCA) that the analyst can use as a starting point when identifying attributes. The list is not exhaustive, but represents the general kinds of attributes to be considered by the analyst when identifying attributes for a task.

Similar to tasks, attributes be dependent. In Table 3-2, for example, lethality is dependent on weapon system accuracy. Accurate delivery of a weapon will help enable it to produce lethal effects against a target. Understanding these interdependencies is critical to identifying potential capability tradeoffs of significance in the study.

Once the attributes are identified, the analyst identifies the operational conditions associated with each task. Operational conditions are described in terms of factors and descriptors. A factor is a variable of the environment that affects task performance. A descriptor is a set level within the range of the factor. In Table 3-1, terrain is an example of a factor with two descriptors, mountainous and plateau. It is important to understand and address the key factors and associated descriptors that influence performance of tasks.

When identifying operational conditions, the analyst must consider the operational context defined in the study. Operational context is a fundamental part of CBAs, pre-MDD analyses, and AoAs since it provides a common frame of reference that covers the full spectrum of relevant operational situations. Operational context includes descriptions of various operational elements such as scenarios, vignettes, locations, physical environments, enemy order of battle, and threats. It is important that the analyst

understand the operational context and ensure the operational conditions defined for the measures align with the operational context used in the study.

3.3. Step 2: Develop Measures Associated with the Task Attributes

Once the tasks, attributes, and conditions are identified, the analyst can proceed with developing the measures. Recall that a measure conveys information about the dimensions, capacity, or amount of an attribute. As is the case for tasks, attributes, and conditions, the analyst should use the CBA(s) and capability requirements document(s) as sources of information for developing measures for the pre- MDD analysis and AoA. Expert elicitation and brainstorming are also useful for gathering information needed to develop measures. For the CBA, the analyst primarily uses findings and data from previous studies, expert elicitation, and brainstorming to develop measures (see the CBA Guidebook, Vol 2c for more information).

When developing measures, the analyst must focus on the attributes associated with the tasks. For each attribute, there is at least one measure, and perhaps several, are required to measure the attribute. As shown in Table 3-2 below, the task "Strike Target" has three attributes associated with it. Two of the attributes, accuracy and timeliness, require multiple measures.

Table 3-2: Multiple Attributes and Measures Example

Task	Attribute	Measure
Strike Target	Accuracy	Miss Distance
		Impact Angle Error
		Impact Heading Error
	Timeliness	Time to Launch
		Time to Strike
	Lethality	Probability of Kill

Measures should address what is most important in accomplishing the tasks. The focus is on the operational effect and the attributes supporting or enabling the operational effect. In most studies, a combination of different types of measures (MOEs, MOSs, and MOPs) are used as shown in Figure 3-2. Figure 3-2 is an example of a measure dendritic for an AoA study that shows the relationships between measures and tasks as well as the numbers and types of measures used in the study. The measure dendritic is used to highlight critical or essential aspects of a capability (measures with KPP or KSA labels). These measures serve as the basis for developing recommendations for potential KPPs and KSAs and their associated threshold and objective values.

As shown in Figure 3-2, there may be cases when supporting measures are appropriate, although there is no requirement to have supporting measures. Supporting measures are used to highlight high- interest aspects of a parent measure or provide a causal explanation of a parent measure. For example, a parent measure (probability of kill) could have circular error probable as a supporting measure.

Probability of kill is likely to be affected by the accuracy (circular error probable) of the weapon as well as other factors such as weapon yield and blast fragmentation pattern. By measuring circular error probable and using it as a supporting measure, the analyst can provide more insights about the kill performance of a weapon.

A parent measure may have one or more supporting measures that may be MOEs, MOSs, or MOPs. Each parent measure and supporting measure should have its own metric, criteria, and data. Figure 3-2 shows three examples of parent measures with supporting measures.

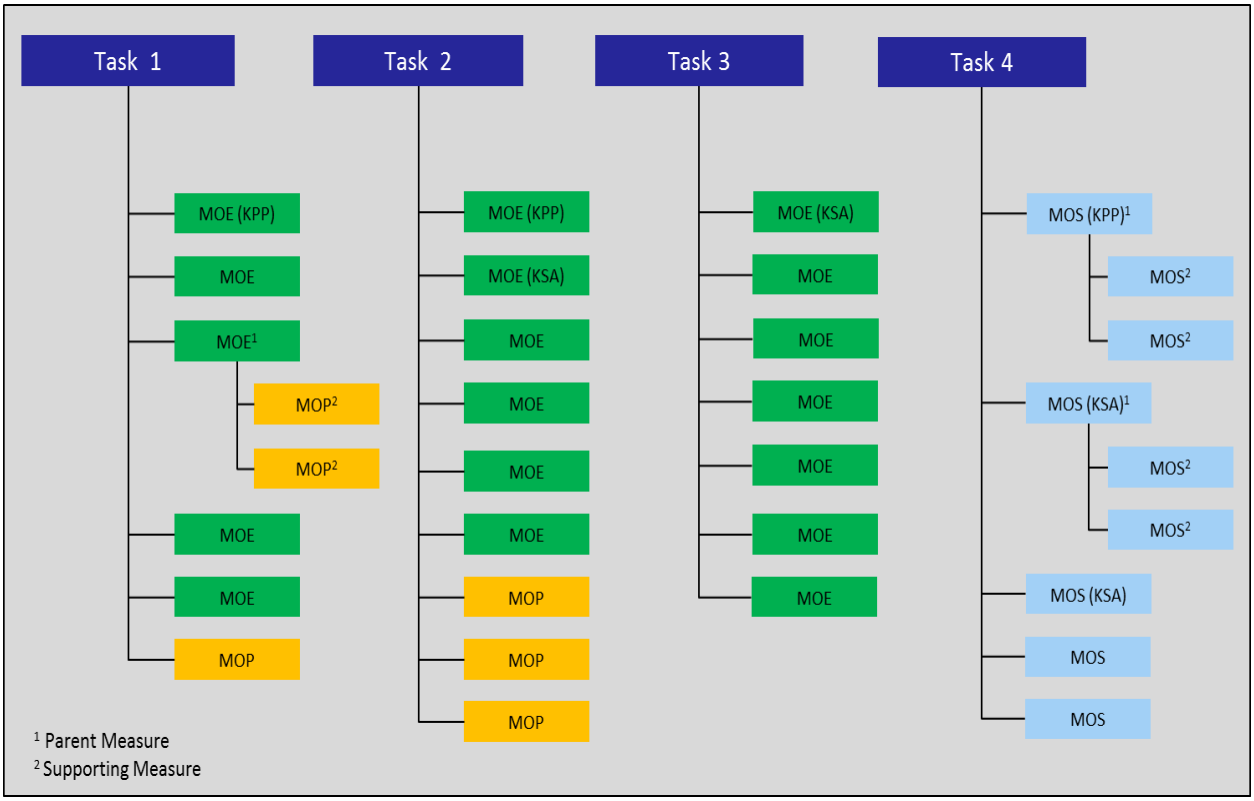


Figure 3-2: Measure Dendritic Example

Note on the Measure Dendritic

The "keep it simple" principle applies when developing the structure of the measure dendritic. With the study purpose and study questions in mind, the analyst should design a dendritic structure that is simple as possible, but suitable to meet the objectives of the study. The number of tasks and measures as well as parent/supporting measure structures add complexity and should be scrutinized to ensure the identified tasks, measures, and measure structures are absolutely necessary.

There are several best practices the analyst should follow when developing measures. First, it is important to keep the measures as simple as possible. If it is too difficult to establish criteria for a measure, then the measure should be rewritten. Second, measures should not be listed more than once for a task, although a measure may be used for more than one task. Third, a measure should not be used as an umbrella or placeholder measure to roll-up or summarize supporting or other measures. Finally, the metric, criteria, and conditions should be treated as separate elements that are associated with the measure and not stated in the measure itself. The analyst should refer to Section 3.5 (Step 4: Verify the Measures) for basic guidelines to follow when developing measures.

3.4. Step 3: Identify Measure Criteria and Metrics

As noted previously, user-established criteria are criteria that are explicitly stated or implied in a capability requirements document (Initial Capabilities Document (ICD), Capability Development Document (CDD), and CDD Update). The analyst should review these documents, if developed, to identify user-established criteria that are relevant to the measures assessed in the study.

When user-established criteria do not exist, criteria are developed to enable the analyst to assess the measure. These criteria are "identified criteria". Sources of information that can be used to develop identified criteria include CONOPS, CONEMPs, Tactics, Techniques, & Procedures (TTPs), previous CBAs, AoAs, and other studies. Along with these sources of information, the analyst can use expert elicitation with appropriate subject matter experts to develop identified criteria.

The analyst must document the source and rationale for measure criteria (both user-established and identified). This is especially important for identified criteria since the criteria have not been previously defined in a capability requirements document.

Note on Identified Criteria

The analyst must obtain user concurrence of identified criteria that have been developed for a study. User concurrence will help mitigate any credibility concerns that may arise later in the study.

In determining the appropriate metric to use for a measure, the analyst will likely require input from subject matter experts to understand what is important in the measurement. Whether the mean, 99th percentile, maximum, or minimum should be used as a metric will depend on the capability needed by the user or warfighter. Subject matter experts knowledgeable of the area of interest can help the analyst determine the appropriate metric for each measure. Finally, the analyst should check to ensure the units of the metric match the criteria values.

Selecting the right metric for a measure also requires an understanding of the data that will be collected. Statistics such as mode, mean, median, and percentage require different mathematical computations and produce values that can vary significantly due to characteristics of the data. Data characteristics such as skewness and variability, for example, can significantly affect metric computations. In these cases, some metrics may not be appropriate since they can affect the meaning of data by hiding information or producing misleading results. An understanding of the data is essential to determining the appropriate metric for a measure.

Though there are many metrics to choose from, the most commonly used metrics in CBAs, pre-MDD analyses, and AoAs are shown in Table 3-3. The table provides a description of each metric as well as the data collection and analysis methods that are typically used to produce and analyze data for the metric. Depending on the data that is produced, some data collection and analysis methods may have limitations on what metrics can be used. For example, expert elicitation and survey research typically produce nominal and ordinal data. For these types of data, the mode and median are customarily used, whereas the mean would not be appropriate.

Table 3-3: Examples of Commonly Used Metrics

Metric	Description	Data Collection/Analysis Association
Mode	Measurement that occurs with greatest frequency in the data set.	Expert elicitation; survey research (questionnaire)
Median	When the number of measurements is odd, it is the middle number when the measurements are arranged in ascending or descending order. When the number of measurements is even, it is the mean of the middle two numbers.	Expert elicitation; survey research (questionnaire)
Minimum	Smallest measurement in a data set.	Modeling and simulation; parametric analysis
Maximum	Largest measurement in a data set.	Modeling and simulation; parametric analysis
Mean	Equal to the sum of measurements divided by the number of measurements contained in a data set.	Modeling and simulation; parametric analysis
Ratio	A comparison or relationship between two numbers or quantities. Ratios can be shown using the ":" to separate values, or as a single number by dividing one value by the other value.	Expert elicitation; survey research (questionnaire)
Proportion	A fraction of the total that possesses a certain attribute.	Expert elicitation; survey research (questionnaire)
Range	Measure of dispersion or spread. Typically expressed as minimum and maximum values.	Modeling and simulation; parametric analysis
Percentage	A number, ratio, or proportion expressed as a fraction of 100. It is often denoted using the percent sign (%).	Expert elicitation; survey research (questionnaire); modeling and simulation; parametric analysis
Percentiles	The values that divide a rank-ordered set of elements from the smallest to the largest into 100 equal parts.	Expert elicitation; survey research (questionnaire); modeling and simulation; parametric analysis
Probability	A measure of the likelihood that the event will occur. The probability of any event can range from 0 to 1.	Modeling and simulation

It is important that the analyst be aware of the tendency by other team members and subject matter experts supporting the study to confuse metrics with units of measurement. Measurement is defined as the assignment of numerals to objects or events according to rules (Kerlinger, 1986). A unit of measurement is a quantity used as a standard of measurement. Some examples of units of measurement

are shown in Table 3-4. When developing metrics for measures, the analyst should be attentive to this issue and, if necessary, provide clarification of the terms.

Table 3-4: Units of Measurement Examples

Variable of Interest	Units of Measurement
Time	Nanoseconds, milliseconds, seconds, minutes, hours, days, months, years
Distance	Inches, feet, miles, nautical miles, fathoms, furlongs
Weight	Ounces, pounds, tons, grams, kilograms
Volume	Ounces, pints, quarts, gallons, milliliters, liters, cubic inches, cubic feet
Height	Inches, feet, millimeters, meters, kilometers
Speed	Miles per hour, kilometers per hour, knots per hour
Altitude	Above ground level, mean sea level
Area	Square inches, square feet, 463L pallet positions
Concentration	Particles per liter, agent-containing particles per liter of air

3.5. Step 4: Verify the Measures

The last step in the process entails verifying the measures. This requires checking each measure to ensure it is stated properly and is relevant to the task and attribute. The remainder of this section provides some basic guidelines to help the analyst in verifying the measures for a study.

3.5.1. Write the Measure Statement Without Referencing the Metric

When possible, write measure statements without referencing the metric. It is important to note that there are cases when it is impractical to write a measure statement without referencing the metric (e.g., probability of kill, probability of survival). The measure statement conveys information about the attribute that will be measured. In the example below, time to deliver the message is the element of interest associated with the timeliness attribute. The first measure statement example references the metric (percentage) and obscures the time element of interest in the measure statement. The measure statement is correctly written in the second example which addresses the time element of interest and maintains the metric as a separate element associated with the measure statement.

Incorrect			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Percentage of messages delivered	Percent	> 95% within 3 minutes
Correct			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Time to deliver message	95th Percentile	< 3 minutes

3.5.2. Write the measure statement without referencing the criteria

The measure statement should not contain the criteria as shown in the first measure example that follows, although criteria will be used to assess the measure. Time to deliver the message is the element of interest and the criteria (within 3 minutes) will be used to rate the measure. The second measure example addresses the criteria as a separate element associated with the measure statement.

Incorrect			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Percentage of messages delivered within 3 minutes	Percent	> 95%
Correct			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Time to deliver message	95th Percentile	< 3 minutes

3.5.3. Write the measure statement without referencing the conditions

The measure statement should not describe the conditions of the measurement as shown in the first measure example below. Address the conditions as a separate element associated with the measure as shown in the second measure example.

Incorrect			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Percentage of messages delivered within 3 minutes in contested environments	Percent	> 95%
Correct			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Time to deliver message	95th Percentile	< 3 minutes
Conditions: permissive and contested environments			

3.5.4. Use a Measure Only Once per Task

A measure should not be listed more than once for a task. In the first measure example that follows, MOEs 3 and 4 are the same measure, though the measurement will be taken under different conditions. As shown in the second measure example, one measure is stated and the measurements will be taken under two different conditions.

Incorrect			
Task 1: Provide Situational Awareness			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	MOE 3: Time to deliver message in permissive environment	95th Percentile	< 3 minutes
	MOE 4: Time to deliver message in contested environment	95th Percentile	< 5 minutes
Correct			
Task 1: Provide Situational Awareness			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	MOE 3: Time to deliver message	95th Percentile	< 3 minutes (permissive environment)
		95th Percentile	< 5 minutes (contested environment)
Conditions:	Permissive and contested environments		

3.5.5. Do Not Use a Measure as an Umbrella or Placeholder

It is not appropriate to use a measure as an umbrella or placeholder for rolling-up or summarizing supporting or other measures as shown in the first measure example below. Each measure should have its own metric, criteria, and data. In the second measure example, the parent measure can be assessed separately since it has its own metric, criteria, and data rather than basing the assessment on the outcome of the supporting measures. The supporting measures are designed to provide additional insights into key system performance characteristics that help enable survivability.

Incorrect			
Attribute	Parent Measure	Metric	Criteria (threshold)
Survivability	Survivability	Measure rating based on the lowest rating of the supporting measures	
	Supporting Measures	Metric	Criteria (threshold)
	Number of threat emitters detected	Percentage	> 95%
	Number of threat emitters identified	Percentage	> 95%
	Number of threat emitters jammed	Percentage	> 95%
Correct			
Attribute	Parent Measure	Metric	Criteria (threshold)
Survivability	Probability of survival	Probability	> .85
	Supporting Measures	Metric	Criteria (threshold)
	Number of threat emitters detected	Percentage	> 95%
	Number of threat emitters identified	Percentage	> 95%
	Number of threat emitters jammed	Percentage	> 95%

3.5.6. Consider Measuring Levels of Performance

When appropriate, the analyst should consider developing measures that distinguish between multiple levels of performance. Such measures provide more information about the true performance of a system or entity. When the underlying performance is binomial, there are only two possible outcomes (e.g., yes or no, pass or fail). The fuse of a munition, for example, either works or not. If the underlying performance is not binomial, then it is possible to develop measures that capture multiple levels of performance.

As shown in first measure example below, two levels of performance are being measured: messages delivered in five minutes or less, and messages delivered in over five minutes. Since the underlying performance is not binomial, it is possible to measure additional levels of performance as shown in the second measure example.

Measuring Two-Levels of Performance			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Time to deliver message	95th Percentile	≤ 5 minutes
Measuring Multiple Levels of Performance			
Attribute	Measure	Metric	Criteria (threshold)
Timeliness	Time to deliver message	95th Percentile	≤ 5 minutes
		75th Percentile	≤ 3 minutes
		50th Percentile	≤ 1 minute

4.0. MEASURE ANALYSIS AND RATING

4.1. Measures Framework

The analyst develops a measures framework to describe the attributes and measures associated with each task and the data collection and analysis methods that will be used in the study. The measures framework is useful for informing the study team, stakeholders, and study oversight groups of the key elements of each measure and analysis methods that will be used in the study. The example shown in Table 4-1 is a measures framework for a notional aircraft electronic warfare system.

There are many methods the analyst can use to collect data and information needed to analyze measures. For each measure, the analyst must consider various factors when selecting the appropriate data collection and analysis method(s). Typically, the analyst must use several different methods to address all the measures in a study. In the example shown in Table 4-1, four different analysis methods were selected. The data collection method chosen by the analyst is important since the data collected will dictate the analysis methods that can be used. For example, data collection methods that produce qualitative data (nominal or ordinal) have limitations on what analytical techniques can be used.

In determining the appropriate data collection and analysis methods, the analyst must understand the capabilities and limitations of the methods. This is particularly important when determining how operational conditions associated with the measures will be addressed. For example, if the threat environment is an important operational condition for a measure, selecting a method that does not enable the analyst to address the threat environment would not be appropriate. In cases when operational conditions cannot be fully addressed, the analyst must document them as limitations of the study.

In some situations, the analyst must use different data collection and analysis methods for an individual measure. This is necessary when the systems, concepts, or alternatives being assessed in the study are at different levels of maturity and definition. For example, alternatives categorized as non- developmental are likely to be very well-defined and have significant amounts of data for specific measures that can be analyzed parametrically or through M&S. In contrast, alternatives that are categorized as developmental may have less definition and data, requiring the analyst to use other methods such as expert elicitation to collect and analyze data for the same measures.

Regardless of the data source and analysis approach used, the analyst must have confidence in the data to make inferences and draw conclusions from the results. Furthermore, the analyst must ensure the distinction between empirical data and expert judgment data is maintained by clearly identifying which analyses are based on empirical data and which are based on expert judgment data.

Table 4-1: Measures Framework Example

Task	Attribute	Measure	Metric	Criteria	Analysis
Enhance survivability	Survivability	Probability of survival	Probability	≥ .85	M&S (BRAWLER)
	Conditions:	Combat range (beyond and within threat detection range); engagement environment (contested, highly contested)			
Detect and identify threats	Completeness	Number of threat detections	Percentage	≥ 98% of threats	Parametric analysis
	Accuracy	Number of threat identifications	Percentage	≥ 95% unambiguous identification of threats	Parametric analysis
	Conditions:	Electronic signal density (high); emitter environment (red, blue, grey, and white); threat classes (low to high priority)			
Sustain and maintain	Availability	Operational availability (Ao)	Probability	≥ .98	M&S (LCOM) ; Expert elicitation
	Reliability	Weapon system reliability	Probability	≥ .98	Comparative analysis; Expert elicitation
	Conditions:	Operations tempo (peacetime, wartime)			

4.2. Data Collection and Analysis Methods

This section provides an overview of data collection and analysis methods that can be used to assess measures developed for the CBA, pre-MDD analysis, and AoA. The data collection and analysis methods described in this section are the most commonly used, so it is not meant to be a comprehensive discussion of all possible methods.

4.2.1. Literature Review

The literature review is useful for creating a foundation to demonstrate knowledge of the current state of the field and should be conducted for most, if not all, studies. Through literature reviews, the analyst can integrate sources of information to identify patterns and determine what is known and what, if anything, is missing. Literature reviews enable the analyst to compare and contrast methods, approaches, and findings and critically discuss the strengths and weaknesses of the sources. By reviewing related studies, the analyst can learn how particular measures were developed and used in the analysis. This will enable the analyst to determine whether specific measurement scales, data collection methods, and analysis techniques can be applied in the study. Finally, the literature review can complement other data gathering techniques such as expert elicitation, brainstorming, and modeling and simulation.

The analyst should consider various sources of information and data such as published and unpublished studies, reports, and papers. Findings and data from previous studies and reports in the area of interest are excellent sources to use in the CBA, pre-MDD analysis, and AoA. In addition, MAJCOMs typically have

SharePoint sites and other repositories of information that may be relevant to the area of interest in the study. Annex E provides a list of frequently used databases of government-sponsored technical documents.

4.2.2. Expert Elicitation

Expert elicitation is a structured method of gathering expert judgment and answering questions concerning issues or problems of interest in a study. Expert elicitation is a form of survey research that can be used to gather a variety of data and information associated with measures such as tasks, attributes, conditions, measure criteria, and measure values. The Delphi method, developed by the RAND Corporation in the 1950s, is one of the first recognized expert elicitation methods. Over the years, many other elicitation methods have been developed and used by various organizations in both the private and public sectors.

There is a variety of terms used to describe expert judgment such as expert opinion, subject matter expert assessment, subject matter expert analysis, subjective judgment, and expert knowledge.

Whatever it is called, expert judgment is the data given by an expert in response to a question and represents an expression of opinion based on knowledge and experience. Judgment is shaped by the expert's state of knowledge at the time of the response to the question. And because experts have different experiences and knowledge, their judgments can differ and change over time as new information is learned.

Since expert judgment is affected by the approach used to gather it, a specially designed process is required that includes procedures for developing questions, conducting the elicitation, and handling biases that may arise. Once the questions have been developed, the analyst uses personal or group interviews to conduct the elicitation. Personal interviews are usually done in private and in person and allow the interviewer to gather in-depth data from the experts without distraction or influence by other experts. Group interviews are conducted in person through a structured approach that defines when and how experts express and discuss their opinions. Although the process is formal and structured, it can differ in terms of the degree of interaction between experts, level of detail in information elicited, number of meetings, type of communication mode, and degree of structure in the elicitation process.

When analyzing responses collected through expert elicitation, the analyst can mathematically aggregate the responses using simple algorithms such as the mean and median regardless of whether the responses were elicited from experts separately or in a group. For example, if experts are asked to provide an estimate of a system's reliability (i.e., a probability value), the analyst can use the mean, median, or other simple algorithms to aggregate the estimates. More complex weighted means can be used to give more weight to experts who are viewed as having more expertise, although the prevailing recommendation among practitioners in expert elicitation is to use equal weights since it is a simple and robust method for aggregating responses. It is important to note that measurement scales such as the Likert scale produce ordinal data, so it is important to use appropriate statistics such as the mode or median.

For additional information on expert elicitation, please see Annex F. Information about developing measurement scales for questions used in expert elicitation can be found in Annex G.

4.2.3. Survey Research (Questionnaire)

Another form of survey research entails administering a questionnaire to gather data for analysis. The analyst can administer the questionnaire either electronically or in paper form. Whatever form is used, good questions and proper administration are essential to collecting meaningful data.

Good questions are unmistakably clear, precise, and unambiguous and ensure the recorded responses align with what the analyst is trying to measure. Questions are specifically worded to avoid creating different interpretations of what is being asked. Differences in answers should be due to differences among respondents rather than from different interpretations of the question's wording. If respondents do not have the same understanding of what the question asks for, error is likely to result. Good questions are both reliable (i.e., provide consistent responses in comparable situations) and valid (i.e., answers correspond to what they are intended to measure).

Crafting good questions requires careful forethought and a sound approach. Subject matter experts (e.g., aircraft operators, logisticians, intelligence experts) who are not selected to be respondents can assist in developing the questions as well as any assumptions, definitions, or other supporting information associated with the questions. Expert insights gleaned during the question development process will help ensure the questions are collecting the information of interest in the study. The CBA, pre-MDD analysis, and AoA typically require many different types of experts, so it is critical to have the right ones participating at the right time.

The process entails drafting a set of initial questions and using a small group of experts to design the final questions. Feedback from experts will be helpful in determining how specific questions should be worded, order and number of questions, and question format. Pre-testing the questions with several other experts can help refine the questions and identify problems such as unclear wording or misreading that must be addressed prior to using the questions in the survey.

There are several aspects of questions that should be considered during the question development process. For instance, whether a question is open or closed can significantly affect the type of data that is collected. Closed questions provide a list of acceptable responses to the respondent, whereas open questions do not provide the acceptable responses. According to Fowler (1993), respondents perform more reliably in answering closed questions since the responses are given. Furthermore, the analyst can more reliably interpret the meaning of the answers. Open questions are appropriate in situations where the list of possible responses is long, making it impractical to present to the respondents. Responses to open questions describe more closely the real views of the respondents and can elicit unanticipated responses.

Proper questionnaire administration includes providing the instrument to respondents, conducting a quality control check of the responses, ensuring the respondents understand all questionnaire items, and actively investigating reasons for certain responses (generally those that are ambiguous or unexpected). There are several guidelines the analyst should consider when administering the questionnaire:

- Instructions should be clear and brief and question forms should be few in number to reduce respondent confusion
- The number of questions and question wording should be kept to a minimum
- Questions should follow a logical order (e.g., time sequence, process related)
- Questions should be asked in a neutral format without leading statements or clues to desired responses

Questionnaires should never be simply handed to respondents who are then asked to "fill them out and return them whenever you can." A much more effective approach is for the analyst to schedule a specific time and place for the respondents to gather and complete the questionnaires. The analyst remains with

the group to field questions and clarify items that may be confusing or are being misinterpreted. As questionnaires are returned, the analyst should carefully examine them to ensure:

- A response alternative has been selected for all items
- The respondent viewed the scale directions correctly (an indication that this may have occurred is when responses from a respondent are mostly opposite to those of other respondents)
- Responses to open-ended questions and any other comments can be read and understood

If there are issues with any questionnaire responses, the analyst should review and resolve them with the respondent immediately. It is never good practice to put off addressing questionnaire problems to a later date as memories fade and people may become unreachable as they move on to other activities.

For additional information on developing questions, please see *Survey Research Methods* by Fowler listed in Annex B (References and Information Sources). Information about developing measurement scales for questions used in survey research can be found in Annex G.

4.2.4. Brainstorming

Brainstorming is a technique that can be used with a small group (ideally 10 or fewer members, but the nature of the problem might necessitate more) to generate ideas about various aspects of measures such as measure selection, data collection methods, and analysis techniques. It can be conducted in- person or electronically. The main principles include focusing on quantity, withholding criticism, welcoming unusual ideas, and combining and improving ideas. Although there are a variety of techniques, the nominal group and group passing techniques are commonly used:

- Nominal group technique encourages all participants to have an equal say in the process. Participants write down their ideas anonymously and a moderator collects the ideas and presents to the group for a vote. Top ranked ideas are sent back to the group or subgroups for more brainstorming and elaboration.
- Group passing technique entails each person in a group writing down an idea on a piece of paper, then passing the paper to the next person who adds thoughts. This continues until everyone gets his or her original piece of paper back. In the end, each group member will likely have an extensively elaborated idea.

4.2.5. Modeling and Simulation

Modeling and simulation (M&S) can be used to generate data for computing metrics for measures developed in a study. A model is a physical, mathematical, or logical representation of a system, entity, phenomenon, or process and is used when it is impossible or impractical to assess a system, entity, phenomenon, or process in the real world. A simulation is a method for implementing the model over time. M&S selection and development is a systematic and iterative process. Before M&S selection and development can begin, the analyst must first conduct an M&S needs and objectives analysis.

In conducting the needs and objectives analysis, the analyst must develop a prioritized list of measurable needs and objectives. In addition to the analysis capabilities that are required (e.g., system or process characteristics to be modeled or represented, output data to be produced for analyzing measures), the analyst should consider the cost, schedule, and personnel constraints of the study when developing the list of needs and objectives. Like other methods, M&S is used to obtain information to solve a problem and inform a decision, although not every problem requires or even benefits from using M&S. In some

cases, other methods may be cheaper, faster, and still meet the needs and objectives of the study. The decision to use M&S should be determined through careful definition of the study purpose and objectives.

As part of the needs and objectives analysis, the analyst must have an understanding of how the variables will be used in the study. As shown in Figure 4-1, the analyst specifies variables that will be used to represent a system, entity, phenomenon, or process as well as variables that will be used to analyze measures for the assessment.

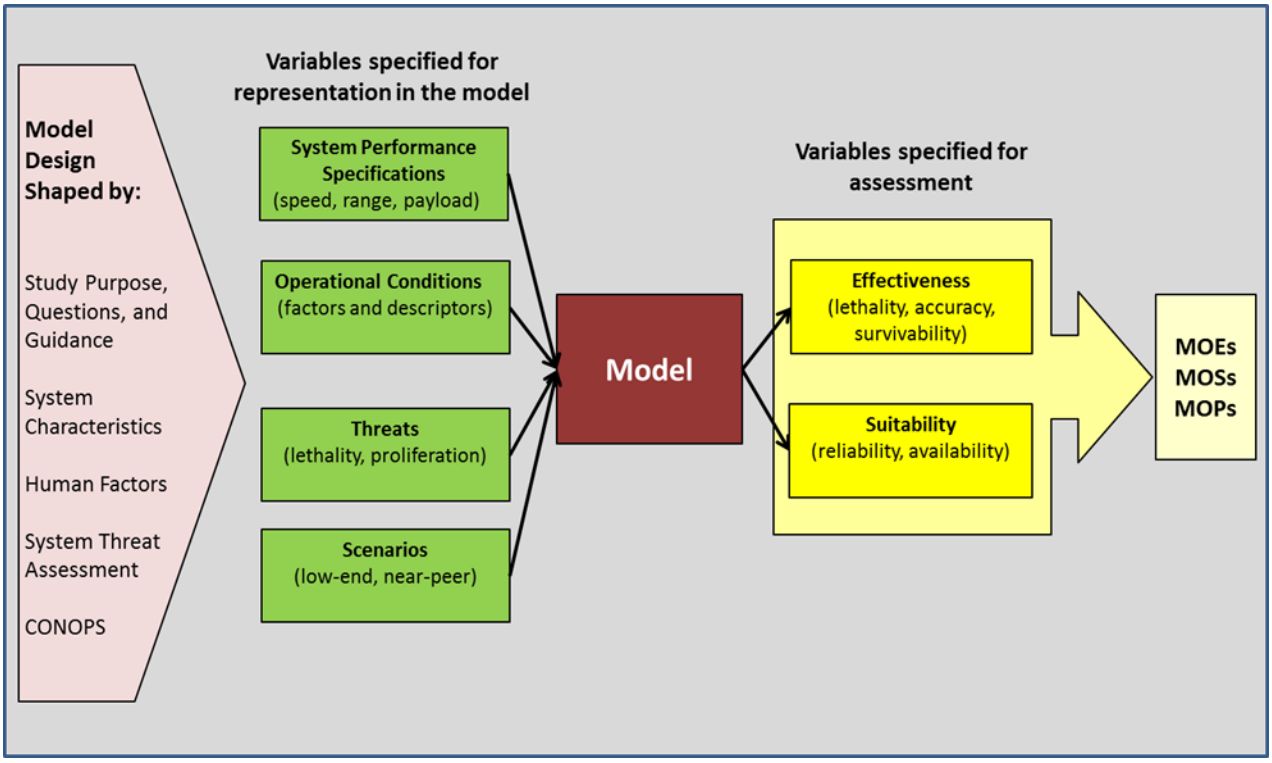


Figure 4-1: Model Variable Relationship

If M&S is the best method, the analyst must determine whether any existing M&S applications are appropriate for the problem. The analyst must examine the capabilities and limitations of the applications, particularly the data input requirements and data output characteristics. It may be necessary for the analyst to evaluate multiple candidates to determine the most appropriate application for the study. Finally, all M&S applications, whether existing and new, must be accredited for use in the study.

If existing M&S applications cannot be reused, then a new M&S application must be developed (see Annex H for more information about M&S development). New M&S application development is more costly and time consuming compared to reusing existing applications, so the overall costs and risks must be considered before proceeding with development. Given the short time frame of the CBA, pre- MDD analysis, and AoA, M&S development must start well before the analysis for which the M&S will be used. Furthermore, new M&S applications must first be verified and validated before they can be accredited for use in the study.

Most M&S applications enable the analyst to use various descriptive and inferential statistics to gain insights about the data. These statistics permit the analyst to identify the points of central tendency,

relationships among variables, and the spread and skewness of data. In addition, the analyst can produce statistics for determining whether measure criteria are met or not.

Additional information regarding M&S can be found at the SAF/SA Studies, Analyses, and Assessments, website listed in Annex B. This website contains descriptions of various M&S applications in the Air Force Standard Analysis Toolkit (AFSAT). AFSAT is an Air Force-approved set of government sponsored computer models and simulations used to conduct analysis in support of decisions spanning the requirements, development, acquisition, and test cycles. Many of the AFSAT M&S applications have been used in previous pre-MDD analyses and AoAs.

4.2.6. Parametric Analysis

The word "parametric" is derived from the word "parameter" which has specific meanings in various fields. A parameter can be generally defined as a measurable factor that can be varied to determine a range of possible outcomes or results. For example, size, weight, and power are parameters that can be varied to produce different physical configurations of a system. Parametric analysis entails using parameters in functions or equations to categorize, order, manipulate, or summarize data.

With an understanding of the parameters of interest in the study, the analyst creates functions or equations that express the relationships among the parameters. Some parameters may be dependent on other parameters, while others may be independent. The analyst must also understand the nature of the output that will be produced from the functions or equations and how it will be used to assess specific measures in the study.

Once the functions or equations are developed, the analyst typically analyzes the data using a spreadsheet or other data processing application. The analyst uses the output that is produced to determine whether measure criteria are met or not.

4.2.7. Comparative Analysis

Comparative analysis is often used in assessing the effectiveness, sustainability, and cost of new systems or concepts. The purpose of the comparative analysis is to select or develop a baseline comparison system that represents characteristics of a new system for estimating parameters and determining effectiveness, sustainability, and cost.

A baseline comparison system may be developed using a composite of elements from different existing systems when a composite most closely represents the design, operation, and support characteristics of a new system. The analysis requires the use of historical data of similar existing systems that are relevant to the system being assessed in the study. If the analyst must estimate parameters of the new system, then current systems which are similar to the new system concept must be identified.

The level of detail required in describing comparative systems will vary depending on the level of detail known about the new system design (e.g., operational and support characteristics) as well as the accuracy required in the estimates. Early in a system life cycle, when the design concept of a new system is very general, only a general level comparative system description can be established by the analyst. For this preliminary analysis, the analyst should identify existing systems and subsystems useful for comparative purposes.

4.2.8. Concept Characterization

A concept is defined as a prospective materiel solution to a capability gap(s) and is described in a Concept Characterization and Technical Description (CCTD) document. CCTDs contain data and information that

the analyst can use when developing and analyzing measures in a study. Examples of information contained in the CCTD include the following:

- Technical planning and analyses that have been accomplished
- Operating environment the concept is expected to be employed in, threats the concept will encounter, and other systems the concept must operate or interface with
- Capabilities needed, required enabling capabilities, operational concepts, mission tasks, key performance parameters and system attributes, and measures
- Concept architectural and design information, supportability and sustainment features
- Description of efforts to develop, test and evaluate, manufacture, and sustain a concept
- DOTmLPF-P implications and interdependencies
- Trade space characterization
- Probability and consequence of risks and mitigation strategies

In some situations, the analyst may be able to directly use data and information from a CCTD for analyzing a measure. For example, to determine whether a system meets a pallet position threshold standard for transportation, the analyst uses the pallet position parameters provided in the CCTD. In these situations, it is important that any data values used in this manner are scrutinized by subject matter experts to help ensure validity of the data. In some cases, it may be necessary for the analyst, based on advice from subject matter experts, to apply an adjustment factor (degradation or augmentation) to the data values provided in the CCTD document before they are used in analyzing the measures. The use of an adjustment factor may be based on various reasons such as knowledge of past performance of a similar concept or accounting for operational conditions that are planned or expected. It is important that the analyst document the rationale used when applying an adjustment factor.

For additional information regarding concepts and CCTD document requirements, see the AF/A5/7 Capability Development Guidebook, Volume 2D, Annex A, Analysis of Alternatives.

4.2.9. Sensitivity Analysis

Sensitivity analysis entails varying parameters to gain insights into performance changes in a concept, system, or alternative of interest in a study. Sensitivity analysis can enhance the credibility of the analysis and help identify significant performance tradeoffs. The main purpose of the sensitivity analysis is to highlight the stability or robustness of a concept, system, or alternative being assessed in a study.

There are several approaches the analyst can take in conducting the sensitivity analysis. One approach involves altering certain assumptions that define performance parameters to identify changes in the results of one or more measures and the operational impacts associated with the changes. For example, varying size, weight, and power parameters based on new assumptions for a system may show significant changes in range and speed performance. Range and speed may be key measures of a system that have the potential to become KPPs or KSAs in a future program. In this case, the sensitivity analysis provides additional insights into the stability of these key measures of performance when assumptions are changed.

Another approach entails altering the operational conditions or scenarios to assess capabilities and limitations of systems in different environments. Using the results of the analysis, the analyst can determine how robust a system is in a wider range of operational conditions and scenarios. Whatever

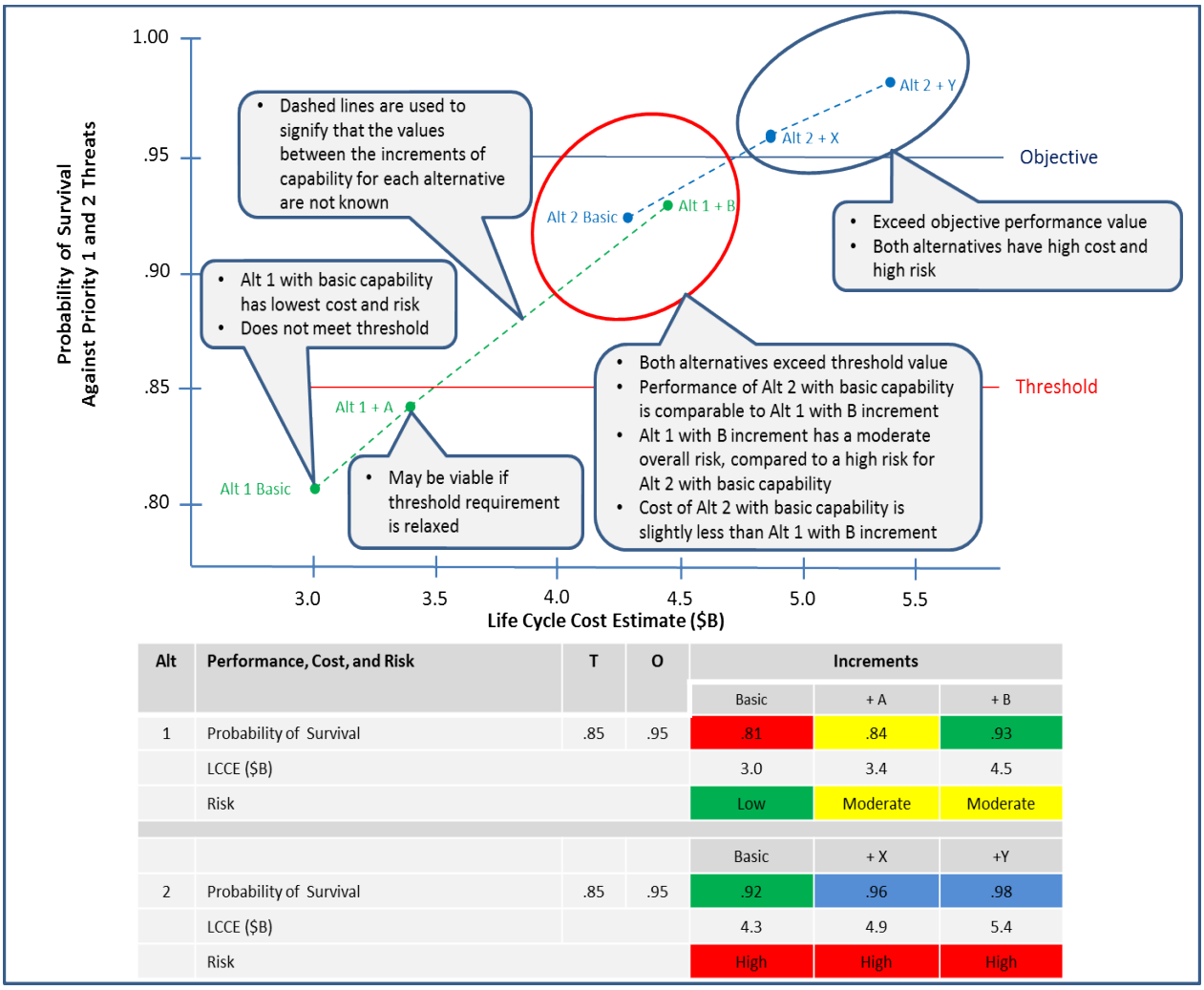
approach is used to conduct the sensitivity analysis, the analysis will enhance the credibility of the study and provide additional insights that will likely be important to the stakeholders and decision-makers.

4.2.10. Cost-Capability Analysis

Cost-capability analysis is used to determine the best-value concept, system, or alternative. Key measures in a study may be used individually or in a composite of two or more measures to describe the capability aspect of the cost-capability tradeoff relationship. The objective of the analysis is to highlight the capability of systems and the associated life-cycle costs. This enables the decision-makers to focus on the tradeoffs between costs and capabilities of the various systems.

Figure 4-2 shows an example presentation of the cost-capability analysis results for a notional aircraft survivability system. Probability of survival was selected since it will be a Key Performance Parameter (other possibilities for the y-axis in this example include reduction in lethality and loss exchange rate). The graphic shows performance against priority 1 and 2 threats which are the most prolific and lethal threats (other possibilities in this example include showing results for one or more scenario vignettes). The life cycle cost estimates (LCCEs) are shown in $B along the x-axis. The table below the graph provides a summary showing the probability of survival, LCCEs, and overall risk ratings of the alternatives. Alternatives 1 and 2 with basic and additional increments of capability are the most viable of the alternatives analyzed in the study and are shown in the figure. Although more sophisticated systems such as alternative 2 with X and Y increments of additional capability (circled in blue) may achieve higher levels of survivability, the costs and risks are high given the leading-edge technology used in the systems. Alternative 2 with basic capability and alternative 1 with the B increment of capability (circled in red) appear to be the best-value alternatives.

The analysis may also show how relaxing a requirement may make other systems more competitive and perhaps more affordable. As shown in the figure, if the probability of survival requirement is slightly reduced, alternative 1 with the A increment of additional capability may be worth considering given its lower cost and moderate risk. The decision-makers must assess the operational impact of a lower probability of survivability requirement and the potential benefits achieved in avoiding costs.



Alt	Performance, Cost, and Risk	T	O	Increments		
				Basic	+ A	+ B
1	Probability of Survival	.85	.95	.81	.84	.93
	LCCE ($B)			3.0	3.4	4.5
	Risk			Low	Moderate	Moderate
				Basic	+ X	+Y
2	Probability of Survival	.85	.95	.92	.96	.98
	LCCE ($B)			4.3	4.9	5.4
	Risk			High	High	High

Figure 4-2: Example of Cost-Capability Analysis for an Aircraft Survivability System

4.3. Rating Measures

There are several approaches the analyst can use to present the results of the measure analysis. One approach entails using a measure rating scale to describe whether or not a measure meets the criteria. For measures that have threshold equals objective (T=O) criteria or have no expressed objective criterion, there are four possible measure ratings as shown in Table 4-2. For these measures, the measure value is rated against the threshold criterion. When a measure value does not meet the threshold criterion (yellow and red rating), operational significance becomes the key consideration.

Answers to the following questions will help the analyst determine the significance of the shortfall:

- How close to the threshold criterion is the measure value?
- What is the consequence or impact on the task and mission if the threshold criterion is missed by a certain amount?
- If the shortfall is only under some operational conditions, what is the significance of the impact?

Whether the shortfall is significant or not ultimately depends on the impact to the task. To determine the impact, the analyst should rely on subject matter experts with the appropriate operational experience and expertise to apply judgment and determine the significance of the shortfall to the task. When a shortfall has only minimal operational impact, the measure should be rated as "did not meet criteria—not a significant shortfall." When the shortfall has a substantial or severe operational impact, the measure should be rated "did not meet criteria—significant shortfall." In both cases, it is important to capture the rationale used to justify the rating. This will enable others to evaluate whether the rationale is credible and defensible.

Table 4-2: Measure Rating Scale

Color Code	Rating
G	Met Criteria
Y	Did Not Meet Criteria—Not a Significant Shortfall
R	Did Not Meet Criteria—Significant Shortfall
	Inconclusive or Not Assessed

When there is insufficient information to assess a measure, it should be rated as "inconclusive." When there is no information to assess a measure, it should be rated as "not assessed."

When an objective criterion is expressed, an alternative rating scale which incorporates an additional rating for the objective criterion is shown in Table 4-3.

Table 4-3: Measure Rating Scale for Measures with Objective Criterion

Color Code	Rating
B	Met Objective
G	Met Threshold
Y	Did Not Meet Threshold—Not a Significant Shortfall
R	Did Not Meet Threshold—Significant Shortfall
	Inconclusive or Not Assessed

4.4. Analysis Pitfalls to Avoid

There are several pitfalls the analyst should avoid when analyzing measures. One such pitfall is the use of measure weighting schemes. Measure weighting schemes can oversimplify results and potentially mask important information. The example shown in Table 4-4 illustrates how measure weighting is dependent on the group determining the weighting and may not be representative of what stakeholders, senior leaders, or decision makers would consider important. The subject matter experts (SMEs) in this example value reducing error (weighted angle and heading error) significantly more than the stakeholders. On the other hand, the stakeholders value minimizing miss distance and time to strike and weighted them significantly more than the SMEs. As a result, the total weighted scores are very different by group. In addition, the stakeholder score masks the poor angle and heading performance since this performance was weighted very low.

Table 4-4: Weighting Measures Example

Measure	Description	Results	SME		Stakeholder	
			Weight	Score	Weight	Score
1.0	Miss Distance	75	.20	15.00	.49	36.75
2.0	Angle Error	9	.35	3.15	.08	0.72
3.0	Heading Error	15	.38	5.70	.06	0.90
4.0	Time to Strike	90	.07	6.30	.37	33.30
Total Weighted Score:				30.15		71.67

When a measure weighting scheme must be used, as in the case when Multi-Objective Decision Analysis (MODA) is used, the analyst must ensure the weighting values are developed by a group of experts with the appropriate experience and expertise. OAS recommends a group of at least six experts to ensure that a diversity of opinions is considered. Although using one or a few experts may require less effort on behalf of the analyst, using a group of experts will lend much more credibility to the weighting values.

Ideally, the group should provide a range of possible weighting values for each measure or item (e.g., a task) being weighted to enable the analyst to conduct a sensitivity analysis later in the study. The weighting values should be reviewed and approved by the sponsor, study oversight group, stakeholders, or decision-makers involved in the study. Finally, the analyst should conduct a sensitivity analysis to demonstrate the stability of the results to changes in weighting values within the specified ranges associated with the measures or items.

Another pitfall to avoid is the inappropriate analysis of ordinal data. Recall that ordinal data is one of the types of data that is usually collected through questionnaires or interview questions. In the example shown in Table 4-5, data for the measure was collected through a questionnaire and exhibits the properties of ordinal data (i.e., data are assigned numbers such that the order of the numbers reflects an order relation based on the attribute). The analyst, however, incorrectly selected the mean as the metric. Despite six of the seven respondents agreeing that a two-level maintenance concept can be used for the system, the measure is rated “did not meet criteria” because 3.8 is not greater than or equal to the threshold value (4).

Table 4-5: Inappropriate Analysis of Ordinal Data Example

Measure Description

Task	Attribute	Measure	Metric	Criteria (Threshold)	Data
Maintain and Sustain System	Maintainability	Logistician rating of maintainability (MOS)	Mean	≥ 4	Logistician responses to questionnaire item with 5-point Likert scale (see below)

Questionnaire Item and Responses

A two-level maintenance concept can be used to maintain this system.					
	Strongly Disagree 1	Somewhat Disagree 2	Neither Agree or Disagree 3	Somewhat Agree 4	Strongly Agree 5
Number of Responses:	0	0	1	6	0
Mean Response: 3.8 ((1 x 3) + ((6 x 4))/7					

To rectify this situation, the analyst selects the mode as the metric (Table 4-6). Using the mode as the metric, the analyst rates the measure as "met criteria" because the mode value of 4 corresponds to "somewhat agree." This example highlights how using inappropriate metrics can sometimes produce different results.

Table 4-6: Appropriate Analysis of Ordinal Data Example

Measure Description

Task	Attribute	Measure	Metric	Criteria (Threshold)	Data
Maintain and Sustain System	Maintainability	Logistician rating of maintainability (MOS)	Mode	≥ 4	Logistician responses to questionnaire item with 5-point Likert scale (see below)

Questionnaire Item and Responses

A two-level maintenance concept can be used to maintain this system.					
	Strongly Disagree 1	Somewhat Disagree 2	Neither Agree or Disagree 3	Somewhat Agree 4	Strongly Agree 5
Number of Responses:	0	0	1	6	0
Mode Response: 4 (Somewhat Agree)					

5.0. INTERPRETING AND REPORTING RESULTS

This chapter provides information that will help the analyst interpret and report measure results. It describes fundamental principles and guidelines to good writing that will help enhance the quality of a report. There is also a discussion of how the analyst uses the results of the measure analysis to assess the overall operational impact at the task and mission levels and the extent to which the capability gaps have been mitigated.

5.1. General Reporting Principles and Guidelines

There is a tendency to give inadequate attention and effort to reporting results and conclusions. The unfortunate consequence of this practice is that a poorly written final report or presentation can significantly diminish the value of a study. Although the analysis may be brilliant, most readers or listeners will be influenced by the quality of the reporting. Tufte (1997) describes how the clarity and excellence in thinking is linked to the clarity and excellence in the display of data: "When principles of design replicate principles of thought, the act of arranging information becomes an act of insight."

Given the importance of a good presentation, the analyst has a special obligation to clearly and objectively communicate the results of the study. Fortunately, there are guidelines the analyst can follow to effectively present study results. Emory (1985) offers the following to enhance the quality of a report:

- Prewriting considerations. Before writing, there are several factors the analyst should consider. Foremost, the analyst should keep the purpose of the study in mind when reporting results. Studies are initiated to achieve specific objectives and address questions from stakeholders and decision-makers. Keeping the study purpose in mind will help the analyst focus on meeting the objectives of the study and answering the study questions. Another factor the analyst should consider is who will read the report. Understanding the needs and biases of the readers will help the analyst determine the discussion length and level of detail that will be required. The greater the gap in knowledge of the subject between the reader and analyst, the greater the challenge for the analyst to fully explain the findings.
- Writing outline. A writing outline helps specify what to write and how to state it. By using a writing outline, the analyst can express the essential thoughts associated with a specific topic. Below is an example of a writing outline for reporting measure results:

  A. Measure statement

  Criteria and criteria reference or rationale

  Measure rating

  Measure rating discussion

  Rationale or justification for rating

  Task and mission performance implications

- Presentation considerations. Good presentation is essential to conveying information clearly and accurately. The following are fundamental guidelines to good writing that will help enhance the quality of a report:

- Choose words that communicate thoughts fully, clearly, and accurately. Plain discourse not only helps enhance readability and comprehensibility, but also avoids ambiguity. Jargon or arcane words do not facilitate understanding and should not be used.
- Summarize and repeat critical or difficult points to ensure the reader gains an understanding of the message. Tables and graphics are also useful for explaining critical or difficult points.
- Use a topic sentence to capture the main thought or subject of the paragraph. A topic sentence helps prepare the reader for the rest of the paragraph and provides a focal point for the supporting details, facts, figures, and examples.
- Use shorter paragraphs to highlight key points and provide a visual relief to readers. Avoid using large blocks of unbroken text since it produces a daunting appearance that is unpleasant to readers. Each paragraph should represent a distinct thought. As a general rule, a paragraph longer than half a page should be scrutinized to ensure it is necessary.
- Use headings and subheadings to create homogeneous sections of the report. Headings and sub-headings help organize the report and serve as signs for the reader to follow.
- Indent parts of text that represent lists or examples.
- Use table and figure labels that are self-explanatory.
- Proofread the document for incorrect spelling, poor punctuation, and improper grammar. Proofreading, preferably by several people, is essential to catching these mistakes and making the necessary corrections (if possible, a review by a professional technical editor can help enhance the quality of the report as well).

There are many references the analyst can use to facilitate good writing. Two examples include the Air Force's *Tongue and Quill* and the American Psychological Association's *Publication Manual*. Some general principles and guidelines from these publications include the following:

- Active/passive voice. There is a tendency to overuse the passive voice in technical writing. Although passive voice is sometimes appropriate (i.e., when the doer or actor of the action is unknown, unimportant, obvious, or better left unnamed), the analyst can enhance the quality of the report by using active voice. Active voice maintains the natural subject-verb-object pattern and conveys the message more clearly and concisely with fewer words. As a general rule, to identify passive voice the analyst should watch for forms of the verb "to be" (am, is, are, was, were, be, being, been) and a main verb usually ending in "ed" or "en." There is also a tendency to confuse passive voice with past tense. Past tense (along with present tense and future tense) is a tense of a verb and is not the same as passive voice. Below is an example of a sentence written in active and passive voice (note the subject-verb-object pattern of the active voice):

  Passive: The ball was thrown by the girl.

  Active: The girl threw the ball.

- Fewer words (economy of expression). Short words and sentences are easier to understand than long ones. The longer it takes to say something, the weaker the communication. Unnecessary words do not help convey a message to the reader and should be removed or replaced with working words. Each word in a sentence should be checked to determine whether the message changes when the word is removed from the sentence. As a general rule, sentences more than

20 words should be examined to determine whether the message can be conveyed more effectively with fewer words or by dividing the sentence into multiple shorter sentences.

- Orderly presentation. The analyst should aim for continuity of words, sentences, and paragraphs from the opening statement to the conclusion. Continuity can be achieved through punctuation marks and transitional words. Punctuation marks cue the reader to pauses (comma, semicolon, and colon), stops (period and question mark), and detours (dash, parentheses, and brackets). Transitional words help maintain the flow of thought. Some examples include the following:
  - Time links: then, next, after, while, since
  - Cause and effect links: therefore, consequently, as a result
  - Addition links: in addition, moreover, furthermore, similarly
  - Contrast links: but, conversely, nevertheless, however, although, whereas

5.2. Interpreting Results

Before measure results can be reported, they must first be interpreted by the analyst. More than just presenting the results, interpretation entails making inferences and drawing conclusions from the results of the analysis. Interpretation is the essence of research, requiring the analyst to search the results for meaning and implications. The interpretation is not only within the study, but also in relation to the results of other studies. In the end, the results should speak for themselves. The analyst is simply conveying the message candidly and precisely.

To facilitate an understanding of the data, the analyst should start by determining whether there are any relationships or associations between the variables of interest in the study. There are two basic types of relationships: dependent and independent. A dependent relationship is one in which there are both independent and dependent variables. The variation of one variable (the dependent variable) depends on the variation of one or more independent variables. In an independent relationship, there are two or more variables of interest, but none are dependent on or influenced by the others. There are statistical techniques the analyst can use to identify these relationships. For instance, analyses such as correlation, regression, and discriminate analysis are commonly used to identify dependent relationships, whereas factor analysis can be used to identify independent relationships.

Posing simple questions can also be very helpful to understanding and interpreting results. Some examples of these questions are:

- What is causing the inflection (knee in the curve), plateau, drop, or spike in performance?
- Why does performance change, or not change, when operational conditions change?
- Are there prominent or critical parameters that influence how well a task or mission is achieved?
- What are the parameter interdependencies and what impacts do they have?

The analysis required to finding answers to questions such as these can be very insightful and will certainly enhance the analyst's understanding of the data. With this knowledge, the analyst will be better able to clearly and fully interpret the results of the study.

5.3. Reporting Measure Results

After all the measures have been rated, the focus of the assessment shifts from individual shortfalls at the measure level to the collective operational impact at the task level. The analyst must rely on specific

evidence in the study and operational experience and expertise of subject matter experts to assess the overall impact to a task. The assessment must be defensible and credible since the foremost concern on the skeptical reader's mind is the "so what" question (e.g., What is the relevance of the issue? How important is it? Why should I care?). Since there is seldom one right answer, the quality and weight of evidence is crucial to answering these questions. Through effective communication, decision-makers should ascertain that the results are valid and the assessment is sound and credible.

In some cases, there may be one or more measures that are very influential on how well a task is achieved. Such measures may address prominent attributes or parameters associated with the task and have the potential to become KPPs and KSAs. The analyst should focus the discussion on these measures by explaining the relationships and impacts to task performance.

In other cases, there may be measures that have significant interdependencies that must be considered when determining the significance of the impact. For example, a particular system may exhibit superior performance in detecting threats but performs marginally in identifying threats. Detection and identification are interdependent capabilities and fundamental to the tasks of finding and tracking threats. When explaining the operational impact, it is important that the analyst maintain a holistic view that is based on an understanding of the interdependencies that exist.

The analyst should avoid relying on the preponderance of measure ratings to assess the collective impact at the task level. For instance, stating that three out of five measures met the criteria so the task is assessed as "green" oversimplifies the assessment and can be misleading. In addition, mathematical and heuristic-based rollup or weighting techniques are never the best way to communicate results.

Although simple to use, these techniques can mask important information that underpins the assessment. In cases when there is insufficient information to make an assessment, the analyst should simply state that the results are inconclusive and explain why.

There are several approaches the analyst can use to present the results of the task level assessment. One approach entails using a task rating scale to help describe the impact at the task level. A task rating scale enables the analyst to assign an overall task rating based on the results of the measures that support the task. The task rating scale shown in Table 5-1 is comprised of four color-coded ratings with definitions. When using a rating scale such as this, the analyst should use subject matter experts with relevant experience and expertise to determine the appropriate rating. Given that the ratings are subjectively determined, it is particularly important that the analyst fully explain the rationale used to assign the ratings in the assessment discussion. This will enable readers to ascertain the validity of the ratings. Lastly, the analyst can use other rating scales, but must ensure the scale ratings are sound and the associated rating definitions are clear.

Table 5-1: Example of a Task Rating Scale

Color Code	Rating	Definition
G	No or Minimal Operational Impact	No or some effectiveness and/or sustainability shortfalls identified with minimal impact on the task
Y	Substantial Operational Impact	Effectiveness and/or sustainability shortfalls identified with substantial impact on the task
R	Severe Operational Impact	Effectiveness and/or suitability shortfalls identified with severe impact on the task
	Inconclusive	Insufficient information to support an assessment

Once the tasks have been assessed, the analyst can evaluate the collective operational impact at the mission or higher level, if necessary. At the mission level, the analyst must consider how well each task is achieved and how it impacts mission accomplishment. It is likely that the contribution or influence of each task to mission accomplishment will vary (i.e., some tasks may be more important than others in accomplishing the mission). With assistance from subject matter experts with the appropriate operational experience and expertise, the analyst should address as part of the assessment discussion the overall impact of each task on the mission.

Another aspect the analyst must address is the degree to which the capability gaps have been mitigated and the impact of the associated operational risks. The analyst uses the collective results of the measure analysis, task assessment, and mission or higher-level assessment as well as the operational experience and expertise of appropriate subject matter experts to explain the extent to which the gaps have been mitigated and the impact of the operational risks. Although it is subjective, the assessment must be supported by a credible and defensible explanation. The analyst should focus on the most important influencing aspects of the measures, tasks, and mission or higher level to explain the degree to which the capability gaps have been mitigated and the impact of the associated operational risks.

ANNEX A: Acronyms

AFSAT	**Air Force Standard Analysis Toolkit**
AFTL	**Air Force Task List**
ANOVA	**Analysis of Variance**
AoA	**Analysis of Alternatives**
APA	**Additional Performance Attribute**
CBA	**Capabilities-Based Assessment**
CCTD	**Concept Characterization and Technical Description**
CDD	**Capability Development Document**
CONEMP	**Concept of Employment**
CONOPS	**Concept of Operations**
CONPLAN	**Concept Plan**
CDD Update	**Formerly Capability Production Document**
DAU	**Defense Acquisition University**
DOTmLPF-P	**Doctrine, Organization, Training, materiel, Leadership and Education, Personnel, Facilities, and Policy**
DPS	**Defense Planning Scenario**
ICD	**Initial Capabilities Document**
ISC	**Integrated Security Construct**
JCA	**Joint Capability Area**
JCIDS	**Joint Capabilities Integration and Development System**
JCTD	**Joint Concept Technology Demonstration**
JMETL	**Joint Mission-Essential Task List**
KPP	**Key Performance Parameter**
KSA	**Key System Attribute**
LCCE	**Life Cycle Cost Estimate Modeling and Simulation**
M&S	**Modeling and Simulation**
MAJCOM	**Major Command**
MDD	**Materiel Development Decision**
METL	**Mission-Essential Task List**
MODA	**Multi-Objective Decision Analysis**
MOE	**Measure of Effectiveness**
MOP	**Measure of Performance**
MOS	**Measure of Suitability**
OAS	**Office of Aerospace Studies**
OPLAN	**Operation Plan**
OSA	**Other System Attribute**
SME	**Subject Matter Expert**
TTPs	**Tactics, Techniques, and Procedures**
UJTL	**Universal Joint Task List**

ANNEX B: References and Information Sources

Government Documents

AFI16-1001, Verification, Validation and Accreditation (VV&A), 29 Apr 2020.

AF/A5/7 Capability Development Guidebook, Volume 2D, Annex A, Analysis of Alternatives

AF/A5/7 Capability Development Guidebook, Volume 2C, Capabilities Based Assessment

JCIDS Manual

Web Links

AF/A5DR - Requirements Policy & Integration:

https://www.my.af.mil/gcss-af/USAF/ep/globalTab.do?channelPageId=s6925EC1352150FB5E044080020E329A9

SAF/SA – Studies and Analyses, and Assessments Analytical Community: https://www.my.af.mil/gcss-af/USAF/ep/browse.do?programId=t6925EC2F6AD90FB5E044080020E329A9&channelPageId=s6925EC13500D0FB5E044080020E329A9

Air Force e-Publishing Online: http://www.e-publishing.af.mil/

Defense Technical Information Center: www.dtic.mil

DAU Glossary: https://www.dau.edu/glossary

Professional Publications

Ayyub, Bilal M. (2001). *A Practical Guide on Conducting Expert-Opinion Elicitation of Probabilities and Consequences for the Corps Facilities*. Alexandria, VA: Institute for Water Resources Report 01-R-01, United States Army Corps of Engineers.

Chan, Gabriel, Laura D. Anadon, Melissa Chan, Audrey Lee. (2001). Expert Elicitation of Cost, Performance, and R&D Budgets for Coal Power with Carbon Capture and Sequestration. *Energy Procedia*.

Churchill, Jr., Gilbert A. (1979). *Marketing Research: Methodological Foundations, Second Edition*. Hinsdale, Illinois: The Dryden Press.

Cooke, Roger M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* New York, NY: Oxford University Press.

Cooke, Roger M. and Katherine N. Probst. (2006). *Highlights of the Expert Judgment Policy Symposium and Technical Workshop*. Washington DC.

DeVellis, Robert F. (1991). *Scale Development: Theory and Applications*. Newbury Park, CA: Sage Publications.

Emory, C. William. (1985). *Business Research Methods, Third Edition*. Homewood, IL: Richard D. Irwin, Inc.

Fowler, Jr., John J. (1993). *Survey Research Methods, Second Edition*. Newbury Park, CA: Sage Publications.

Kerlinger, Fred N. (1986). *Foundations of Behavioral Research, Third Edition*. Fort Worth, Texas: Harcourt Brace Jovanovich College Publishers.

Leedy, Paul D. (1997). *Practical Research: Planning and Design, Sixth Edition*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Meyer, M. A. and J.M. Booker. (2001). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics, American Statistical Association.

O'Brien, Thomas G. and Samuel G. Charlton (editors). (1996). *Handbook of Human Factors Testing and Evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. (2001). *Uncertain Judgements: Eliciting Experts Probabilities*. West Sussex, England: John Wiley and Sons, Ltd.

Publication Manual of the American Psychological Association, Fourth Edition. (1994). Washington DC: American Psychological Association.

Tufte, Edward R. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.

Tull, Donald S. and Del I. Hawkins. (1980). *Marketing Research: Measurement and Method*. New York, New York: Macmillan Publishing Co., Inc.

Zikmund, William G. (1991). *Exploring Marketing Research, Fourth Edition*. Hinsdale, Illinois: The Dryden Press.

ANNEX C: Glossary

Additional Performance Attribute – performance attributes that are not important enough to be considered Key Performance Parameters or Key System Attributes, but still appropriate to include in the Capability Development Document and Capability Production Document. (JCIDS Manual)

Attribute – a quality or feature of something. Attributes of tasks (e.g., survivability, persistence, availability, accuracy, etc.) form the basis for identifying and drafting measures.

Baseline – the capability that currently exists or is programmed for in the future. The Baseline can be Materiel, non-materiel, or a combination of both.

Capability – the ability to complete a task or execute a course of action under specified conditions and level of performance. (JCIDS Manual)

Capability Gap (or Gap) – the inability to meet or exceed a capability requirement, resulting in an associated operational risk until closed or mitigated. The gap may be the result of no fielded capability, lack of proficiency or sufficiency in a fielded capability solution, or the need to replace a fielded capability solution to prevent a future gap. (JCIDS Manual)

Capability Requirement – CRs are Measures of Effectiveness (MOE) in the form of mission focused task statements that are best written in "task, condition and standard" format. CRs are described in relation to tasks, conditions, and standards IAW the Universal Joint Task List or equivalent DoD Component Task List and are thought of as "what needs to be done (the metric), and to what level (the initial value)". If a CR is not satisfied by a capability solution, then there is an associated capability gap. A requirement is considered to be 'draft' or 'proposed' until validated by the appropriate authority. (JCIDS Manual)

Conditions (Operational) – describes the environment under which the mission will be performed.

Criteria (also referred to as Standards) – define the acceptable levels or standards of performance for a metric and are often expressed as a minimum acceptable level of performance (threshold) and desired acceptable level of performance (objective).

Data – individual measurements that are used to compute the metric for a measure.

Key Performance Parameter – performance attributes of a system considered critical or essential to the development of an effective military capability. (JCIDS Manual)

Key System Attribute – performance attributes considered important to achieving a balanced solution/approach to a system, but not critical enough to be designated a KPP. (JCIDS Manual)

Measure – a device designed to convey information about an entity being addressed. It is the dimensions, capacity, or amount of an attribute an entity possesses. - Measure of Effectiveness – a measure designed to correspond to accomplishment of mission objectives and achievement of desired results. (DAU Glossary)

Measure of Performance – a measure of a system's performance expressed as speed, payload, range, time on station, frequency, or other distinctly quantifiable performance features. (DAU Glossary)

Measure of Suitability – a measure of an item's ability to be supported in its intended operational environment. (DAU Glossary)

Metric – a unit of measure that coincides with a specific method, procedure, or analysis (e.g., function or algorithm). Examples include: mean, median, mode, percentage, and percentile.

Objective – an operationally significant increment above the threshold. An objective value may be the same as the threshold value when an operationally significant increment above the threshold is not identifiable.

Other System Attribute – other attributes not previously identified as a KPP, KSA, or APA. Other System Attributes tend to be attributes associated with design, life cycle cost, or risk drivers. (JCIDS Manual)

Stakeholder – any agency, Service, or organization with a vested interest (a stake) in the outcome of the analysis. (AoA Guidebook, Vol 2D, annex A)

Task – describes what is expected to be performed and is commonly expressed as an action or activity.

Threshold – a minimum acceptable operational value of a system capability or characteristic below which the utility of the system becomes questionable.

ANNEX D: Operational Attributes by Joint Capability Area

Examples are provided in the JCIDS Manual 2021, Annex A, Appendix B, Enclosure C

Force Integration Attributes: Accuracy, Adaptability, Comprehensiveness, Credibility, Integration, Timeliness.

Building Partnership Attributes: Agility, Breadth, Depth, Effect, Flexibility, Persistence, Utility.

Battlespace Awareness Attributes: Accuracy, Adaptability, Comprehensiveness, Credibility, Innovativeness, Integration, Interoperability, Persistence, Survivability, Timeliness.

Force Application Attributes: Accuracy, Adaptability, Capacity, Flexibility, Mobility, Persistence, Scalability, Security, Survivability, Timeliness.

Logistics Attributes: Accountability, Agility, Attainability, Capacity, Economy, Effectiveness, Enduring, Expeditionary, Flexibility, Integrated, Networked, Persistence, Precision, Reliability, Responsiveness, Scalability, Simplicity, Survivability, Sustainability, Tailorability, Visibility, Velocity.

Command and Control Attributes: Accessibility, Assured, Agility, Completeness, Interoperability, Latency, Relevance, Reliable, Resilient, Security, Diverse, Simplicity, Timeliness, Understanding.

Communications and Computers Attributes: Accessibility, Accuracy, Agility, Availability, Capacity, Completeness, Controllability, Expeditionary, Flexibility, Integration, Interoperability, Latency, Maintainability, Configurability, Relevance, Reliability, Responsiveness, Robustness, Scalability, Security, Survivability, Throughput, Timeliness, Visibility.

Protection Attributes: Capacity, Effectiveness, Integration, Networkability, Persistence, Responsiveness, Survivability, Speed, Maneuverability, Detectability, Vulnerability, Durability, Resiliency, Recoverability.

Corporate Management and Support Attributes: Accessibility, Accuracy, Auditability, Availability, Efficiency, Integration, Interoperability, Latency, Reliability, Responsiveness, Security, Throughput, Timeliness, Usability, Visibility.

ANNEX E: Literature Review Sources of Information

AcqNotes: https://acqnotes.com/subjects-index

Adaptive Acquisition Framework: https://aaf.dau.edu/

DTIC: www.dtic.mil

Information and Resource Support System (IRSS): https://irss.milcloud.smil.mil/

- Requires SIPRNet Air Force Portal account, and permission from AF/A5DR

Rand Corp: www.rand.org

The Knowledge Management/Decision Support system (KM/DS): For instructions go to SIPR Intelink: https://www.intelink.sgov.gov/my.policy

Manual for the Operation of the Joint Capabilities Integration and Development System

ANNEX F: Using Expert Elicitation in the CBA, Pre-MDD Analysis, and AoA

Introduction

Expert elicitation is a structured method of gathering expert judgment and answering questions concerning issues or problems of interest in a study. The Delphi method, developed by the RAND Corporation in the 1950s, was one of the first recognized expert elicitation methods. Over the years, many other elicitation methods have been developed and used by various organizations in both the private and public sectors. There are numerous examples of its use by federal agencies to include the United States Army Corps of Engineers, Nuclear Regulatory Commission, National Aeronautics and Space Administration, Department of Transportation, Department of Energy, Department of Agriculture, and the Environmental Protection Agency.

Since expert judgment is affected by the approach used to gather it, a specially designed process is required that includes procedures for developing questions, conducting the elicitation, and handling biases that may arise. The process is designed to facilitate thinking and encourage experts to state their true opinions. Through the elicitation process, experts derive judgments from the available body of evidence ranging from direct empirical data to theory. Although the process is formal and structured, it can differ in terms of the degree of interaction between experts, level of detail in information elicited, number of meetings, type of communication mode, and degree of structure in the elicitation process.

Expert elicitation is different from sampling methods since respondents are not considered to be representative of a population (Chan et al, 2010). Instead, respondents are viewed as representing a large body of knowledge. Expert elicitation seeks to reflect the range of credible opinion regarding a specific question or problem, so the foremost concern is the quality and diversity of the participating experts.

After a brief overview of expert elicitation and judgment, this annex presents an approach to conducting expert elicitation in the CBA, pre-MDD analysis, and AoA. It provides insights regarding the selection of experts, development of questions, and design and conduct of the elicitation process.

What is an Expert?

Meyer and Booker (2001) define an expert as "a person who has background in the subject area and is recognized by his or her peers or those conducting the study as qualified to answer questions." It is natural to think of experts as professionals such as scientists, physicians, and engineers, but any person with sufficient knowledge of the subject matter can be considered an expert for the purpose of the study. Although an individual's knowledge is important, other factors such as personality, experience, and expertise in organizing and using his or her knowledge are critical to the success of the elicitation (O'Hagan et al, 2006, p. 27). Achieving a balanced and broad spectrum of viewpoints may require eliciting judgments from individuals with various backgrounds and degrees of expertise.

Expert Judgment

There is a variety of terms used to describe expert judgment such as expert opinion, subject matter expert assessment, subject matter expert analysis, subjective judgment, and expert knowledge. Whatever it is called, expert judgment is the data given by an expert in response to a question and represents an expression of opinion based on knowledge and experience. Judgment is shaped by the expert's state of knowledge at the time of the response to the question, and because experts have different experiences and knowledge, their judgments can differ and change over time as new information is learned.

Expert judgment is commonly expressed in quantitative terms, although it is possible to obtain expert judgment in a variety of other non-numeric or qualitative forms. Some examples of information elicited from experts are shown in Table F-1.

Table F-1: Examples of Information Elicited from Experts

Quantitative	Qualitative
Probability of an occurrence of an event	Impact of a change
Probability of failure of a system	Risks and consequence of a decision
Estimates of ranges of uncertainty	Variables, assumptions, and data used in an analysis
Likelihood of a causal relationship	Elements needed for decision making
Allocation of funding	Failure causes, potential failures, and potential solutions
Rating of the performance of a model	Methods to optimize performance

An Expert Elicitation Approach for the CBA, Pre-MDD Analysis, and AoA

It is necessary to follow a formal and structured process to ensure the information elicited from experts is suitable for analysis. The following describes a seven-step approach to conducting expert elicitation in the CBA, pre-MDD analysis, or AoA. It provides guidelines for the selection and preparation of experts, development of questions, design and conduct of the elicitation process, and analysis and reporting of data.

Step 1. Identify the Need for Expert Elicitation

In conducting the CBA, pre-MDD analysis, or AoA, the analyst must typically deal with many unknowns associated with new and complex concepts. Choosing the appropriate research methods to collect and analyze data is a foremost concern. Study objectives, data accessibility, time and resource constraints, and available tools and techniques are some important factors that the analyst must consider when determining which research methods to use.

Expert elicitation can be a very useful technique for gathering data given the breadth of information that may be collected. Expert elicitation is appropriate in situations where traditional research methods are not feasible or data is insufficient, unattainable, or too costly or impractical to collect. Some examples of the information that can be elicited from experts in these studies include the following:

- Establishing study ground rules, constraints, and assumptions
- Identifying and rating risks and consequences
- Identifying criteria (threshold and objective values) of performance measures
- Providing estimates of performance measures

Step 2. Develop the Questions

Expert elicitation relies on surveys to collect data of some aspect for analysis. Expert judgment is primarily elicited through face-to-face interviews. The choice of whether to use personal interviews (i.e., interview one expert at a time) or group interviews (i.e., interview experts in a group) will depend on various factors such as time constraints and the availability of experts. Whatever method is chosen, using good questions is an essential part of the survey process.

Good questions are unmistakably clear, precise, and unambiguous and ensure the recorded responses align with what the analyst is trying to measure. Questions are specifically worded to avoid creating different interpretations of what is being asked. Differences in answers should be due to differences among respondents rather than from different interpretations of the question's wording. If respondents

do not have the same understanding of what the question asks for, error is likely to result. Good questions are both reliable (i.e., provide consistent responses in comparable situations) and valid (i.e., answers correspond to what they are intended to measure).

Crafting good questions requires careful forethought and a sound approach. Subject matter experts who are not among the experts in the panel can assist in developing the questions as well as any assumptions, definitions, or other supporting information associated with the questions. Expert insights gleaned during the question development process will help ensure the questions are eliciting the information of interest in the study. The CBA, pre-MDD analysis, and AoA typically require many different types of experts (e.g., aircraft operators, logisticians, intelligence experts), so it is critical to have the right ones participating at the right time.

The process entails drafting a set of initial questions and using a small group of experts to design the final questions. Feedback from experts will be helpful in determining how specific questions should be worded, order and number of questions, and question format. Pre-testing the questions with several other experts can help refine the questions and identify problems such as unclear wording or misreading that must be addressed prior to using the questions in the elicitation.

There are several aspects of questions that should be considered during the question development process. For instance, whether a question is open or closed can significantly affect the type of data that is collected. Closed questions provide a list of acceptable responses to the respondent, whereas open questions do not provide the acceptable responses. For closed questions, respondents can perform more reliably in answering the question since the responses are given and analysts can more reliably interpret the meaning of the answers (Fowler, 1993, p. 82). Open questions are appropriate in situations where the list of possible responses is long, making it impractical to present to the respondents. Responses to open questions describe more closely the real views of the respondents and can elicit unanticipated responses.

Whether personal or group interviews are used, there are several guidelines to consider when administering the questions:

- Instructions should be clear and brief and question forms should be few in number to reduce respondent confusion
- The number of questions and question wording should be kept to a minimum
- Questions should follow a logical order (e.g., time sequence, process related)
- Questions should be asked in a neutral format without leading statements or clues to desired responses

Step 3. Select the Experts

Selection criteria define the set of individuals that have a chance of being selected to participate as expert panel members in the study. It is important to establish selection criteria through careful deliberation since the selection of experts is a critical step in the process. Since the expert panel selection is not random, there is a risk of researcher bias when the researcher makes selections based on inappropriate criteria. Selection error present in an expert panel depends on the degree of expertise of the person making the selection decision. It is advantageous to consider a range of possible criteria by drawing from the expertise of the study director, study team members, study advisory group, and other appropriate groups and organizations.

A "good" expert has technical knowledge, experience, and intuition as well as an ability to integrate information and draw conclusions. Criteria such as level of training, type of skill, and years of experience can be used to ensure the panel consists of experts with the proper knowledge and expertise.

Ultimately, selection criteria will depend on the objectives of the study. Table F-2 provides some examples of criteria that can be used to identify experts for participation in a study.

Table F-2: Examples of Selection Criteria

Criteria	Description
Knowledge of Area of Interest	Understanding of the area of interest, reputation as a technical authority, awards received, membership in organizations or groups in the area of interest.
Background and Experience	Years of experience, level and diversity of experience, type and number of past positions held.
Education and Training	Specialized training, type of advanced academic degree(s), special certification(s) and qualifications.
Published Work	Number and quality of publications in the area of interest.
Personal Skills	Interpersonal skills, communication skills, flexibility, impartiality, ability to generalize and simplify.
Economic or Personal Stake	Lack of economic or personal stake in the potential findings.
Availability and Willingness	Availability and willingness to commit the necessary time and effort to participate in the study, willingness to prepare for discussions and provide opinions.

Like other studies, the number of experts used in the CBA, pre-MDD analysis, and AoA will be driven mostly by resources and time available to conduct the study as well as the number and availability of individuals who have the expertise in the area being studied. Although there are no absolute rules regarding the number of experts, large panels increase the likelihood that all possible expert views are represented. While all are knowledgeable of the area of interest, experts have different experiences and perspectives that will shape their responses. Large panels can often produce insights that may not be possible with small panels.

Despite the lack of definitive approaches to determining the appropriate number of experts, a panel of practitioners in expert elicitation recommends at least six experts should be included and that the benefit of including additional experts beyond 12 begins to diminish (Cooke and Probst, 2006, p. 16). Using panels with less than six members will likely reduce the chances of collecting a diversity of information.

Step 4. Prepare the Experts

Once the experts have been identified and selected, the next step entails preparing them for the elicitation by providing relevant information about the study. Experts must have a thorough understanding of the issues before they are ready to answer questions. Issue familiarization is the process used to help the experts understand the issues of interest in the study, purpose of their participation, expectations, study objectives, elicitation process, list of questions, terminology, and key assumptions and definitions. Depending on the objectives of the elicitation, information about the technical aspects of the baseline capabilities, potential solutions, study methodology, and performance measures may be required.

Whether done in a group or individually, it is important to present the same information to ensure a common understanding of the issues. Presentations, briefing books, and other documents should be assembled to provide the relevant information.

Step 5. Conduct the Elicitation

The approaches used to elicit judgments vary widely and will rely to a large degree on the objectives of the study. The amount of time required for the elicitation may range from a few hours to as much as a week depending on the size and complexity of the study. The analyst should consider a number of factors in designing the elicitation:

- Time and resources available for the study
- Type of information to be elicited
- Number of experts
- Amount of time experts will need to provide judgments
- Degree of interaction among the experts
- Number and type of questions
- Format for the answers
- Mode(s) of communication

Type of interview - Expert judgment is elicited through personal or group interviews. Personal interviews are usually done in private and in person and allow the interviewer to gather in-depth data from the experts without distraction or influence by other experts. Group interviews are conducted in person through a structured approach that defines when and how experts express and discuss their opinions.

Although personal interviews can be used, convening an in-person group meeting to conduct the elicitation has several advantages in the CBA, pre-MDD analysis, and AoA. Most importantly, it provides an opportunity to introduce the issue, review the relevant information, and describe the elicitation purpose and process. It can serve as a forum to answer questions, share information, discuss expectations, describe how the results will be used, and gain feedback on any issues that require further clarification or additional information. The major drawback to group elicitation is the undesirable effects of dominant or vocal participants, something that is avoided by eliciting experts individually through personal interviews (Cooke and Probst, 2006, p. 16).

In group elicitations, there are greater demands of time and effort on the interviewer to structure and facilitate the discussions and interactions amongst the experts. The interviewer is responsible for ensuring the integrity of the elicitation process and its implementation by initiating and maintaining effective discussions. Ayyub (2001, p. 18) recommends using a facilitator or moderator to help create an environment that ensures equity in presenting views and a successful elicitation of opinions and information from each expert.

In these studies, gaining insights into the underlying reasoning or rationale of an expert's response may be as important as the response itself. There are several techniques described by Meyer and Booker (2001) that can be used to interview experts and learn the rationale for a response:

- The verbal report involves instructing the expert to think aloud when answering a question and resembles someone talking to oneself. The technique can be time consuming since it is used on one expert at a time. It is important to note that not all experts are capable of verbalizing all their thoughts for various reasons (e.g., too difficult to articulate, thoughts are automatic or

unconscious).

- The verbal probe entails phrasing questions in a way to minimize influencing the expert's thinking. The technique is a quick means of obtaining information and is suitable for both personal and group interviews.
- The ethnographic technique involves transposing the expert's words into questions. Because the questions are based on the expert's own words, it is a non-biasing form of questioning. The technique can be time consuming and is not suitable for group interviews.

In structuring the elicitation, it is important to understand and anticipate bias that may occur. Bias is a skewing that arises from our personal perceptions and understanding. There are various forms of bias and methods for dealing with them. Table F-3 provides a brief description of seven common forms of bias and when they are likely to occur.

Table F-3: Common Forms of Bias (derived from Meyer and Booker, 2001, p. 133)

Bias	Description
Social Pressure – Data Gatherer	Individuals consciously or unconsciously alter the descriptions of their thoughts to gain acceptance and to be seen in the most positive light possible. Data gatherers can intentionally or unintentionally influence the individual through body language, facial expression, intonation, and word choice. More pronounced in cases when the interviewer uses leading questions.
Social Pressure – Group Think	Social pressure from others in a group induces individuals to alter their responses or silently acquiesce to what they believe will be acceptable to the group. More pronounced when individuals in a group desire to remain as members, are satisfied with the group, and view the group as cohesive.
Wishful Thinking	Individuals' hopes influence their judgment—what individuals think should happen will influence what they think will happen. More pronounced when individuals do not have to explain their reasoning and when individuals are personally involved or would gain from their answers.
Inconsistency	Individuals are inconsistent in solving problems—as experts' thinking evolves over time, their current thoughts or answers may contradict those expressed earlier. More pronounced when: 1. Elicitation sessions are long and individuals forget instructions, definitions, and assumptions, 2. Complicated response forms such as probability distributions and percentiles are causing confusion, 3. Experts are asked to consider too many things and become confused and inconsistent.
Underestimation of Uncertainty	Individuals underestimate the uncertainty in the answers they provide. More pronounced when response forms are probabilities and other quantitative estimates.
Anchoring	Individuals receive additional information but do not adjust from their first impression in answering the question. More pronounced when experts have described their positions orally or in writing and fear losing face if they change their response.
Availability	Individuals do not mention more than one or two considerations in giving their responses which can mean the experts are drawing from data that is easier to recall. More pronounced when the expert does not receive any information from others that could help trigger less accessible data when formulating a response.

Several steps can be taken in designing the elicitation process to help mitigate anticipated bias. For example, to reduce social pressure from the data gatherer, the interviewer can use the verbal report, verbal probe, and/or ethnographic phrasing of questions instead of direct questions that may lead the experts. If complicated response forms such as probability and uncertainty estimates are being elicited, prepare the experts for the elicitation by conducting a training session that describes the fundamental principles of the response form. The training will help eliminate the potential of confusion and underestimation and give the experts an opportunity to rehearse providing responses to sample questions in the appropriate form. Finally, as part of the preparation for the elicitation, it is important to make the experts aware of the forms of bias and why they happen. Although bias cannot be completely eliminated,

experts will not be able to control their own tendencies toward bias without first having a good understanding of it.

While much can be done to design the elicitation to help mitigate bias, the interviewer must still be alert to the occurrences of bias during the elicitation process and make the appropriate adjustments to counter it. For example, if there are inconsistencies in responses, the interviewer should ask the experts to reconsider their responses. If fatigue is a factor, the interviewer can shorten the elicitation sessions or schedule breaks to help preclude potential inconsistencies in responses. In group situations, the interviewer should suspect group think is occurring when no one in the group voices a difference of opinion or the experts defer to one or more other experts.

There are many different approaches to interview experts that would be appropriate in these studies. In group situations, one approach commonly used involves interviewing each expert separately, reviewing the answers in a group, and then providing an opportunity for the experts to revise their responses.

Depending on the objectives of the study, the analyst may only be interested in collecting responses to questions, whereas in other cases, the rationale for the response may be required as well. Following are several examples of elicitation methods for group interview situations:

- Each expert is asked to provide a response to a question as well as rationale for his or her response that includes identification of issues that significantly influenced the response. After providing responses, the panel of experts is given an opportunity to review the results. During the review, each expert discusses the rationale for his or her response while the other panel members are encouraged to ask questions and contribute information. Following the review, the experts are given an opportunity to revise their responses and provide rationale in light of what was learned during the discussion. With the submission of the revised responses, the question is closed and the elicitation process resumes with the next question.
- Each expert is asked to provide an initial response to a question. To avoid social pressure, the individual responses are then displayed anonymously to the panel of experts through an on-screen graphical presentation. The experts are given an opportunity to discuss the results of the presentation. Following the discussion, the experts provide a final response. With the submission of the final response, the question is closed and the elicitation resumes with the next question.
- Questions with associated background information are provided to the panel of experts. To encourage knowledge sharing, the experts are given an opportunity to discuss the questions and information as a group. The interviewer monitors the discussion and responds to any questions from the panel members. If necessary, the interviewer provides additional information to help the panel in understanding the issues. The information may be requested by the panel or the interviewer, through observation, deems the information is needed to facilitate the discussion. When the panel discussion is complete, each expert is asked to provide a response to each of the questions. With the submission of the response, the questions are closed and the elicitation resumes with the next set of questions.

In personal interview situations, experts are interviewed separately in face-to-face meetings or by telephone. If the response requires clarification or there is a desire to collect the rationale for the response, the analyst can use the verbal report, verbal probe, or ethnocentric technique described earlier to gather the information. For example, an analyst can instruct the experts to explain in detail heir thinking process as they respond to the questions (verbal report). The verbal probe and ethnographic technique can be used to clarify responses and/or gain more insights into the rationale for the responses.

The questions used in the elicitation will depend on the objectives of the CBA, pre-MDD analysis, or AoA. Questions can be designed to elicit opinions in a variety of forms such as quantities, uncertainties, relationships, parameters, or events. Following are several examples of information that can be elicited:

- In determining the probability of a system failure, experts are asked to provide a best estimate as well as a degree of uncertainty. The best estimate is expressed as a percentage, although the decimal or ratio can be used as well. This estimate is viewed as the median value where there is a 50% chance that the "true" value will be higher, and a 50% chance the "true" value will be lower. Next, the experts are asked to estimate an upper bound where there is a strong likelihood (95% chance) that the "true" value will be lower than the estimate, and only 5% chance that the "true" value will be higher. In the analysis, these estimates are used as the 50^{th} and 95^{th} percentile values.
- After reviewing technical information of a system, the experts are asked to rate how easily the system can be configured for transport. Each expert is asked to answer a series of questions with five-point Likert scales ranging from "strongly disagree" to "strongly agree" and provide written rationale for his or her response. In the analysis, the mode value is determined for each question and the rationale used by the experts is highlighted in the discussion of the results.
- Experts are given an opportunity to review five models used for predicting performance of a system. Each expert is asked to rate the plausibility of each model using a seven-point scale ranging from "Least Plausible" to "Most Plausible" and provide written rationale for his or her response. In the analysis, the responses from the experts are shown graphically along with the median rating for each model. The results provide a discussion of the median ratings and rationale used by the experts in rating the models.

Step 6. Aggregate the Data

In the CBA, pre-MDD analysis, and AoA, there is typically a requirement to report a single value by combining responses. Whether judgments are elicited from experts separately or in a group in some instances, the analyst can mathematically aggregate the responses using simple algorithms such as the mean and median. For example, if experts were asked to provide an estimate of a system's reliability (i.e., a probability value), the analyst can use the mean, median, or other simple algorithms to aggregate the estimates. More complex weighted means can be used to give more weight to experts who are viewed as having more expertise, although the prevailing recommendation among practitioners in expert elicitation is to use equal weights since it is a simple and robust method for aggregating expert judgments (O'Hagan, 2006, p. 222; Meyer and Booker, 2001, p. 329). Measurement scales such as the Likert scale produce ordinal data, so it is important to use appropriate statistics such as the mode or median.

If the judgments are elicited from experts in a group, another option is to use a behavioral aggregation that requires a convergence or consensus of opinion among the experts through discussion and interaction. A major risk of this approach is the undue influence of dominant participants.

Step 7. Report the Results

Since there is both potential value and danger of using expert judgment, some guidelines are necessary when reporting results derived from expert judgment. Traditional scientific research does not explicitly accommodate the use of opinions as scientific data. Expert opinions are subjective beliefs that may be useful data, but not scientific in the sense that it has been subjected to empirical inquiry and test. It is important to ensure the distinction between empirical data and expert judgment data is maintained by clearly identifying which analyses are based on empirical data and which are based on expert judgment

data. Cooke (1991) recommends that sufficient information should be provided about the data and calculations so that the results can be reproduced by others.

Another important consideration is the generalizability of results. Unlike probability sampling, expert elicitation is unlikely to produce results that are representative of a population since all individuals in the population do not have equal chances of being selected. This means the analyst should not make statistical inferences about a population from the expert judgment data. Expert elicitation does not entail randomly selecting individuals with the intent of making inferences about a population, but rather, individuals are selected based on their knowledge and experience with the intent of drawing conclusions about the existing knowledge base.

Finally, the analyst should provide the names and background information of the experts used in the study in the final report. This will help readers ascertain the credibility of the experts.

Summary

Expert elicitation can be a useful technique for gathering various types of data for analysis in the CBA, pre-MDD analysis, and AoA. Expert elicitation is a formal and structured process that entails the selection of experts, conduct of the elicitation, and analysis of data. The approach described in this annex will help ensure the information elicited from experts is properly collected and suitable for analysis. It provides guidelines for the selection and preparation of experts, development of questions, design and conduct of the elicitation process, and analysis and reporting of data.

Attachment A. References

Ayyub, Bilal M. (2001). *A Practical Guide on Conducting Expert-Opinion Elicitation of Probabilities and Consequences for the Corps Facilities*. Alexandria, VA: Institute for Water Resources Report 01-R-01, United States Army Corps of Engineers.

Chan, Gabriel, Laura D. Anadon, Melissa Chan, Audrey Lee. (2001). Expert Elicitation of Cost, Performance, and R&D Budgets for Coal Power with Carbon Capture and Sequestration. *Energy Procedia*.

Cooke, Roger M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* New York, NY: Oxford University Press.

Cooke, Roger M. and Katherine N. Probst. (2006). *Highlights of the Expert Judgment Policy Symposium and Technical Workshop*. Washington DC.

Fowler, Jr., John J. (1993). *Survey Research Methods*, Second Edition. Newbury Park, CA: Sage Publications.

Meyer, M. A. and J.M. Booker. (2001). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics, American Statistical Association.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. (2001). *Uncertain Judgements: Eliciting Experts Probabilities*. West Sussex, England: John Wiley and Sons, Ltd.

ANNEX G: Scale Development

Surveys are used to collect data of some aspect for analysis. Expert elicitation, for example, relies on surveys to elicit information from subject matter experts. All surveys, whether conducted through interviews (personal or group) or questionnaires (electronic or paper), rely on using questions to collect information from respondents.

There are two general types of questions that can be used: closed and open. Closed questions provide a list of acceptable responses to the respondent, whereas open questions do not. For closed questions, the analyst must develop a measurement scale to record responses. Measurement scales, usually referred to as scales, enable the analyst to measure aspects of interest in the study.

This annex provides some guidelines for the analyst to follow when developing scales for questions. More detailed information regarding scale development and survey research can be found in the references listed at the end of this annex.

Nominal and Ordinal Scales

Closed questions typically have nominal or ordinal based scales to measure a response. As shown in Table G-1, the nominal scale uses categories (e.g., yes or no) that have no rank order relationship. In contrast, the ordinal scale uses a rank order relationship. Responses to questions are normally treated as nominal or ordinal data based on the scale used. It is important that the analyst know the type of data and the appropriate statistics that can be used (see Section 2.1 Levels of Measurement for more information).

Table G-1: Nominal and Ordinal Based Scales

Nominal Scale	
The solution will enable the nuclear enterprise to accomplish its mission.	
Yes	No

Ordinal Scale					
The solution will enable the nuclear enterprise to accomplish its mission.					
Completely Disagree	Substantially Disagree	Slightly Disagree	Slightly Agree	Substantially Agree	Completely Agree

The Likert scale, developed by sociologist Rensis Likert to measure psychological attitudes in a scientific way, is an ordinal based scale that is commonly used in studies to measure the level of agreement or disagreement. As shown in Table G-2, the scale is bivalent (two-directional) and balanced (i.e., equal number of positive and negative response alternatives) with a neutral middle. The scale has verbal labels that connote evenly spaced graduations of the response alternatives. Five-point response alternative scales are often used, though seven and nine point scales can be used as well.

Table G-2: Example of a Likert Scale

A two-level maintenance concept can be used to maintain this system.				
Strongly Disagree	Somewhat Disagree	Neither Agree or Disagree	Somewhat Agree	Strongly Agree

Scale Considerations

While the question prompts the response, the scale determines the form of the response. The thought and deliberation that goes into crafting good questions applies as well to selecting the appropriate scales to use. A scale must align with the wording used in the question and the intent of the measure. If a question asks if something is useful, for example, a scale that measures usefulness in some way should be used. In addition, the scale should be linked to the measure to ensure the data can be used to make an assessment as to whether the measure criteria are met or not.

The response alternatives used in scales are chosen for consistency, discriminability, and comprehensibility. Response alternatives with these attributes can help avoid nonresponse and response bias. Examples of five, six, and seven point scales with these attributes are shown in Table G-3.

When determining what response alternatives to use, the analyst should consider the following:

- Response alternatives should retain the same directional order for all questions (i.e., low to high, or high to low) to avoid response errors, unless there is a belief that the order will make a difference in the responses selected.
- Balanced scales such as the Likert scale have an equal number of positive and negative response alternatives and tend to produce distributions that are more nearly normal (O'Brien and Charlton, 1996, p. 84).
- Although greater discriminability can be obtained by adding more response alternatives, more than seven response alternatives increases response variability and lowers overall reliability (O'Brien and Charlton, 1996, p. 87).

Neutral Midpoint

Another aspect the analyst must consider is whether to use a neutral midpoint in a scale. Scales without a neutral midpoint force respondents to select a response that departs from true neutrality which can occasionally result in nonresponse. The drawbacks of forcing respondents to make a choice must be carefully weighed against the benefits of obtaining non-neutral responses.

Table G-3: Examples of Five, Six, and Seven Point Scale

Five Point Scales				
Totally Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Totally Adequate
Completely Unacceptable	Somewhat Unacceptable	Borderline	Somewhat Acceptable	Completely Acceptable
Extremely Difficult	Somewhat Difficult	Borderline	Somewhat Easy	Extremely Easy
Strongly Disagree	Somewhat Disagree	Neither Agree or Disagree	Somewhat Agree	Strongly Agree
Undoubtedly Worse	Moderately Worse	The Same	Moderately Better	Undoubtedly Better
Never	Rarely	Now and Then	Often	Always

Six Point Scales					
Totally Inadequate	Very Inadequate	Somewhat Inadequate	Somewhat Adequate	Very Adequate	Totally Adequate
Completely Unacceptable	Largely Unacceptable	Somewhat Unacceptable	Somewhat Acceptable	Largely Acceptable	Completely Acceptable
Extremely Difficult	Moderately Difficult	Somewhat Difficult	Somewhat Easy	Moderately Easy	Extremely Easy
Completely Disagree	Substantially Disagree	Slightly Disagree	Slightly Agree	Substantially Agree	Completely Agree
Undoubtedly Worse	Moderately Worse	Slightly Worse	Slightly Better	Moderately Better	Undoubtedly Better
Never	Very Rarely	Somewhat Rarely	Somewhat Often	Very Often	Always

Seven Point Scales						
Totally Inadequate	Very Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Very Adequate	Totally Adequate
Completely Ineffective	Largely Ineffective	Somewhat Ineffective	Borderline	Somewhat Effective	Largely Effective	Completely Effective
Extremely Difficult	Moderately Difficult	Somewhat Difficult	Borderline	Somewhat Easy	Moderately Easy	Extremely Easy
Undoubtedly Worse	Moderately Worse	Slightly Worse	The Same	Slightly Better	Moderately Better	Undoubtedly Better
Never	Very Rarely	Somewhat Rarely	Borderline	Somewhat Often	Very Often	Always

"Not Applicable" as a Selection

The selection of respondents to participate in a survey not only requires careful planning and preparation, but also a thorough understanding of the respondent qualifications needed to answer the survey questions. Despite the best efforts of the analyst, there may be cases when respondents are asked questions concerning things about which they do not know. One approach to deal with such a possibility is to include "not applicable" as a selection separate from the response alternatives. A "not applicable" selection indicates the respondent did not have adequate knowledge or experience on which to base an answer.

There are two other reasons for including "not applicable" as a selection. First, it will allow the analyst to better ascertain whether item nonresponse was intentional or unintentional since the likelihood that respondents who do not have a basis for an opinion will intentionally not answer the question is low. Second, a "not applicable" selection helps prevent respondents who do not have adequate knowledge or experience on which to base an answer from selecting a neutral response alternative. Table G-4 shows an example of an item with "not applicable" included as a selection.

Table G-4: Item with "Not Applicable" Included as a Selection

How important is the airborne radiation survey reconnaissance map for the ground planning mission?					
Not Important at All	Not So Important	Neutral	Fairly Important	Very Important	Not Applicable

Comments:

When qualified respondents do select a "not applicable" response, the analyst must investigate the reason for the selection to ensure it was not accidental and that the respondent actually had no basis for an opinion. In the report, the analyst should clearly identify these occurrences, describe the reasons for them, provide justification for their removal from the sample population, and clearly document the actual sample size. Because neutral responses like "Neither Agree or Disagree" are not particularly informative, the analyst should also investigate and document the reasons for these response selections. Finally, the analyst should always investigate and document any response anomalies such as outliers (in either direction) and bi-modal distributions.

When analyzing and presenting data for an item, the analyst should exclude "not applicable" selections from the responses of those who are qualified to answer an item. Including "not applicable" selections with all other responses can produce misleading results and lead to incorrect interpretations of the data. Figures G-1 and G-2 show how the response distributions change significantly when the "not applicable" selections are inappropriately included as part of the data set. In Figure G-1 (without "not applicable" selections), the majority of respondents (8 out of 10) who are qualified to answer the question think the reconnaissance map is fairly or very important for the ground planning mission. With the addition of the "not applicable" selections in Figure G-2, it appears there is no longer a majority of respondents who think the reconnaissance map is important for the ground planning mission. The visual image presented in the figure draws attention to the high number of respondents who did not have adequate knowledge or experience on which to base an answer. In this case, interpreting the results becomes more difficult and may lead to faulty conclusions. For instance, one may incorrectly conclude from the figure that less than half of the respondents think the airborne radiation survey reconnaissance map is important for the ground planning mission.

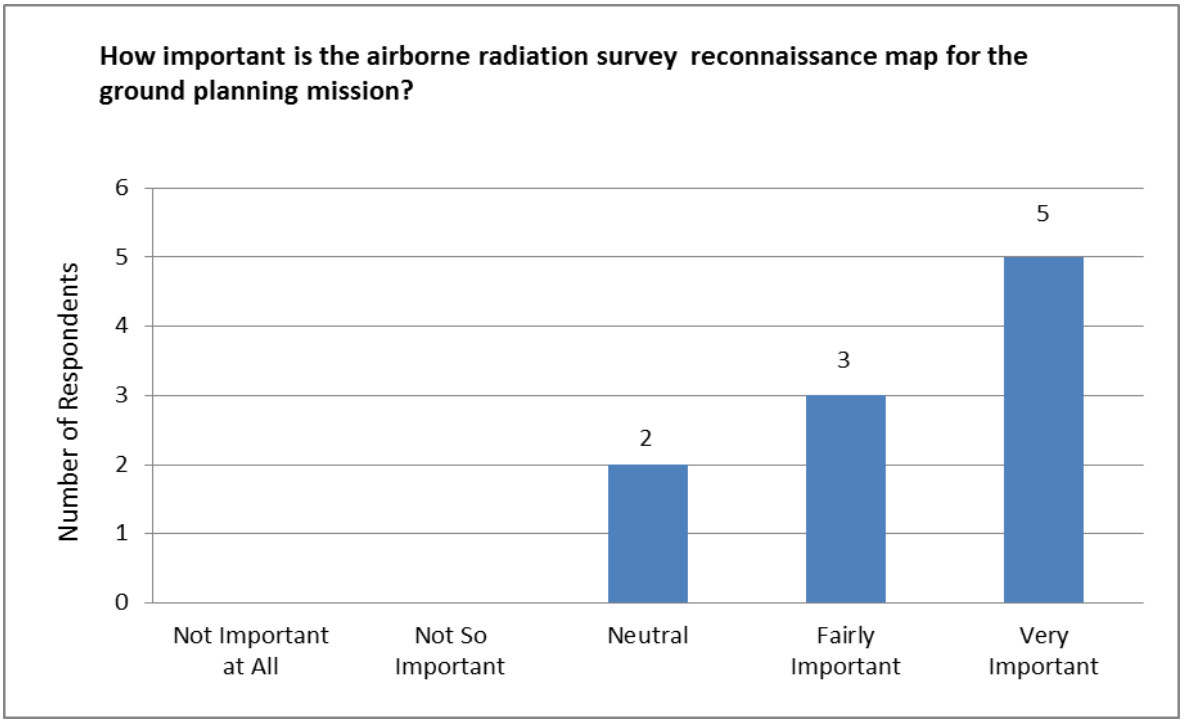


Figure G-1: Bar Chart Display without "Not Applicable" Selections

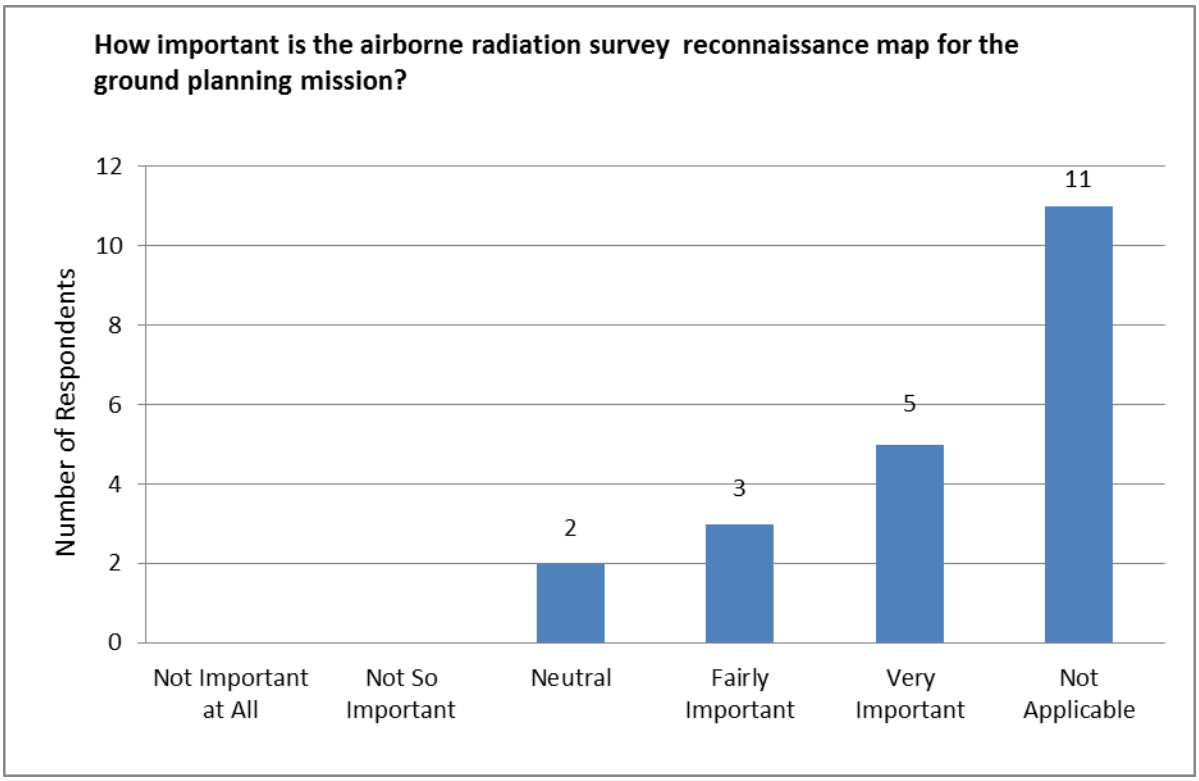


Figure G-2: Bar Chart Display with "Not Applicable" Selections

Attachment A: References

DeVellis, Robert F. (1991). *Scale Development: Theory and Applications*. Newbury Park, CA: Sage Publications.

Fowler, Jr., John J. (1993). *Survey Research Methods, Second Edition*. Newbury Park, CA: Sage Publications.

O'Brien, Thomas G. and Samuel G. Charlton (editors). (1996). *Handbook of Human Factors Testing and Evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc

.

ANNEX H: Modeling and Simulation (M&S) Development Process

This annex provides an overview of the M&S development process. The intent is to provide insights into the tasks, complexity, and level of effort associated with M&S development, rather than a step-by- step approach to developing M&S applications. More detailed information regarding the M&S development process can be found in the references listed at the end of this annex.

Needs and Objectives Analysis

Like other methods, M&S is used to obtain information to solve a problem and inform a decision, although not every problem requires or even benefits from using M&S. In some cases, other methods may be cheaper, faster, and still meet the needs and objectives of the study. The decision to use M&S should be determined through careful analysis and definition of the needs and objectives of the study.

If M&S is the best method, the analyst must determine whether any existing M&S applications are appropriate for the problem. The analyst must examine the capabilities and limitations of the applications, particularly the data input requirements and data output characteristics. It may be necessary for the analyst to evaluate multiple candidates to determine the most appropriate application for the study. Finally, all M&S applications, whether existing and new, must be accredited for use in the study.

If existing M&S applications cannot be reused, then a new M&S application must be developed. New M&S application development is more costly and time consuming compared to reusing existing applications, so the overall costs and risks must be considered before proceeding with development. Given the short time frame of the CBA, pre-MDD analysis, and AoA, M&S development must start well before the analysis for which the M&S will be used. Furthermore, new M&S applications must first be verified and validated before they can be accredited for use in the study.

The needs and objectives analysis begins with developing a prioritized list of measurable needs and objectives. An explicit statement of the M&S needs and objectives is important since it will help enable clear communication throughout the M&S development process. Developing such a statement requires an understanding of how variables of interest will be used in the study. In M&S applications, variables are used to represent a system, entity, phenomenon, or process as well as to produce output data for analysis. In addition to the analysis capabilities that are required, the analyst must consider the cost, schedule, and personnel constraints of the study when developing the list of needs and objectives.

Multi-Disciplinary Team

An essential step in developing a new M&S application is building a multi-disciplinary team chartered to develop the application. This begins with identifying the expertise that is needed and defining the roles and responsibilities of the team members. Defining how team members will interact with each other and how information will be communicated and recorded is essential to fostering mutual understanding and support across the team. The initial M&S development approach should be described in a high-level schedule with milestones, activities, and products to help facilitate understanding of the development effort the team will be undertaking.

Depending on the magnitude of the effort, a number of roles must be filled. Listed below are the most critical roles:

- Sponsor. Identifies need for M&S development, defines M&S requirements, provides resources to develop and implement M&S.
- Program Manager. Plans and organizes resources for M&S development and oversees

preparation of M&S for use, configuration management, and maintenance.

- Developer. Designs and implements the M&S application.
- User. Defines M&S requirements and operates the M&S application.
- Verification and Validation Agent. Performs the verification and validation of the M&S.
- Accreditation Agent. Performs the accreditation of the M&S.

Requirements Analysis

The team conducts the requirements analysis to define specific and accurate requirements for the M&S application to be developed. The overall intent of the requirements analysis is to describe what the M&S application will represent and the level of fidelity that must be achieved (i.e., the accuracy and resolution of the representation). When conducting a requirements analysis, the team should consider the following:

- Requirements should be clear, testable, and trace back to the needs and objectives of the study.
- Requirements should address both representational and operational needs.
  - Representational requirements describe what is represented and how well.
  - Operational requirements describe the conditions and functions that are required by defining the character of the necessary interfaces, computing infrastructure, and control mechanisms.

To define requirements, the team must delineate the scope of the entity being modeled and determine what flows into and out of the entity. This entails defining a conceptual model that describes the inputs, variables, and parameters that will be supplied to the M&S application as well as the outputs that will be produced from the application. The conceptual model is used to transform the M&S requirements into specifications for designing the M&S application. The following are several key questions the team should address when defining the conceptual model:

- What are the constraints, limitations, and assumptions? Funding, personnel, schedule, data?
- What data to include? Physical systems, environment, human element?
- What should be modeled? Equipment, systems, environment, human characteristics, interactions, behaviors?
- What level of detail? System, component, subcomponent? Mission, theater, campaign? Tactical, operational, strategic?

Development Approach

M&S development is a systematic and iterative process that is based on sound systems engineering principles and practices. To define the development approach, the team must consider various factors such as available resources, critical deadlines or milestones, access to programmers and subject matter experts, software support, facilities, and operating system requirements. The team must identify potential tradeoffs associated with these factors since they can influence the design of the final product. For example, to meet a critical milestone, the team may be required to modify system requirements, potentially resulting in less functionality than what was originally planned.

The team should take the time to identify and understand the impact of all possible design options on the effectiveness of the final product. Implementation should not begin until all the impacts are fully understood. Finally, the team should consider conducting verification and validation routinely during the development process.

Implementation and Application

Implementation entails developing and integrating software, acquiring and configuring hardware, and integrating and testing software and hardware. As part of the implementation, verification and validation that is ongoing during development is completed. Verification is the process used to determine whether the M&S application accurately represents the conceptual description and specifications. Validation is the process used to determine whether the M&S application is an accurate representation from the perspective of its intended uses.

Application entails training operators and analysts in the sue of the M&S application and accrediting the application for use in the study. Accreditation is the official certification that the M&S application is acceptable for use for a specific purpose. The accreditation agent relies on the results of the verification and validation as well as other factors to make an accreditation determination.

Attachment A: References and Acronyms

References

AFI 16-1001, Verification, Validation and Accreditation (VV&A), 1 Jun 1996.

Essentials of Modeling and Simulation (https://nmso.navy.mil/ems/welcome.html)

Models and Simulations (M&S) Selection and Accreditation Handbook: A Practical Guide in Support of Analyses of Alternatives, Office of Aerospace Studies, August 2008

Acronyms

ACAT	Acquisition Category
ADM	Acquisition Decision Memorandum
AoA	Analysis of Alternatives
CDD	Capability Development Document
DAE	Defense Acquisition Executive
DoD	Department of Defense
FoS	Family of Systems
ICD	Initial Capabilities Document
IPT	Integrated Product Team
JCIDS	Joint Capabilities Integration and Development System
JROC	Joint Requirements Oversight Council
MDA	Milestone Decision Authority
OAS	Office of Aerospace Studies
OSD	Office of the Secretary of Defense
SME	Subject Matter Expert
SoS	System of Systems

Appendix B - Survey Research

Using Survey Research in the Operational Capability Requirements Study

8 Aug 2023

Preface

In operational capability requirements studies, it is common for the analyst to collect some of the data using one or more forms of survey research. The approach used to gather data influence the data and it is important that the analyst have a good understanding of the fundamental principles of survey research. Good questions and proper administration are essential to collecting meaningful data.

A survey is one of many analytic methods used in the world of operational capability requirements studies. Some problems and questions lend themselves well to surveys, while others do not. The purpose of this handbook is to describe the appropriate uses of survey research and provide insights into planning and conducting survey research in the Capabilities-Based Assessment (CBA), Analysis of Alternatives (AoA), and other types of operational capability requirements studies.

This handbook describes the principles of survey research that ensure questions are both reliable (i.e., provide consistent responses in comparable situations) and valid (i.e., answers correspond to what they are intended to measure). With expert elicitation being a special form of survey research, this handbook also presents an approach to conducting expert elicitation in operational capability requirements studies. This appendix supplements capability development guidebooks by providing a comprehensive discussion of survey research principles.

TABLE OF CONTENTS

1.0. SURVEY RESEARCH

After a brief introduction of survey research and the types of measurement and data, this chapter describes some fundamental principles of designing good questions. A discussion of nominal and ordinal scales and guidelines for developing scales for questions follows. Finally, several principles that the analyst or interviewer[1] should consider when designing, pretesting, and administering a survey are discussed.

1.1. Introduction

Survey research is used to collect data and information of some aspect for analysis. All surveys, whether conducted through interviews (personal or group) or self-administered with questionnaires (electronic or paper), rely on using questions to collect information from respondents.

There are two main reasons for using survey research in operational capability requirements studies.[2] First, an analyst may want to collect someone's opinion or judgment regarding a specific question or issue. For example, an analyst may want to identify the risks associated with each of the alternatives in an AoA study. In this situation, the analyst could conduct a survey with a group of experts to identify the risks. Survey research is designed for collecting an individual's opinion or judgment about something. In other situations, the analyst may want to use survey data to supplement data collected through other methods. In this case, the survey data may be used to corroborate the results in a study.

The second main reason for using survey research is to collect data when it cannot be collected through other methods. For example, it may not be possible to estimate a system's reliability through modeling and simulation, parametric analysis, or comparative analysis. In this situation, survey research may be used to elicit reliability estimates from qualified experts. Survey data is often used to assess measures in the CBA, AoA, or other type of capability requirements study when other data collection methods are impractical or not possible.[3]

1.2. Types of Measurement and Data

As shown in Table 1-1, there are four general levels of measurement: nominal, ordinal, interval, and ratio. The levels of measurement range in sophistication from low (nominal) to high (ratio). Nominal and ordinal levels of measurement are commonly referred to as qualitative measures. Interval and ratio levels of measurement are referred to as quantitative measures. Since the measurement characteristics are different at each level, there are particular statistics that are appropriate for each level.

---

[1] The term "interviewer" refers to the individual who is conducting the survey. The term "analyst" refers to one or more individuals or members of a study team responsible for conducting all or some aspect of an analysis. The interviewer and analyst may be the same individual(s).

[2] Operational capability requirements studies include the Capabilities-Based Assessment, Analysis of Alternatives, and other types of requirements-related studies. Requirements refer to operational requirements that are typically identified during the requirements identification phase of the requirements development process. To annotate this meaning, the term "operational capability requirements" and a shortened version of the term "capability requirements" are used throughout this handbook.

[3] Some examples of measures that may be assessed using survey data include the measure of effectiveness (MOE), measure of suitability (MOS), and measure of performance (MOP). For more information about developing and analyzing measures for the CBA, AoA, or other types of capability requirements studies, please see *The Measures Guidebook, Vol 2k*.

When determining the levels of measurement that will be used, the analyst must consider various factors such as the attribute being measured, purpose of the measurement (e.g., counting objects in categories, attaining a rank order), and data collection requirements and constraints. The analyst should strive to use the highest possible levels of measurement that are suitable for the study. It is important that the analyst has an understanding of the levels of measurement to ensure the appropriate statistics are used.

There are two general types of data: objective and subjective. Objective data is collected with instrumentation or some other direct means. If collected without the use of personal judgment, data collected through direct observation is considered an objective data collection method (e.g., pilot report of airspeed read directly from a display). In contrast, subjective data is based on an individual's judgment or opinion about something. Table 1-2 shows the four possible combinations of measurement and data.

Table 1-1: Levels of Measurement

Category	Level	Description	Appropriate Statistics	
			Descriptive	Inferential
Qualitative	Nominal	Data are assigned the same value or symbol if they have the same attribute. Example: 1 – Male, 2 – Female	Mode, percentages, and frequencies.	Chi-square, binomial test, McNemar test, and Cochran Q test.
	Ordinal	Data are assigned numbers/symbols such that the order of the numbers/ symbols reflects an order relation based on the attribute. Example: 1 – Bad, 2 – Medium, 3 – Good	All statistics permitted for nominal scales plus percentile (e.g., median (50th percentile), 80th percentile, and 95th percentile).	Product-moment correlation, Z-test, T-test, F-test, factor analysis, and ANOVA.
Quantitative	Interval	Data are assigned numbers such that differences between numbers represent equivalent intervals. Example: Temperature in degrees Fahrenheit*	All statistics permitted for ordinal plus mean, standard deviation, and range.	Product-moment correlation, Z-test, T-test, F-test, factor analysis, and ANOVA.
	Ratio	Data are assigned numbers that have all the features of interval measurement as well as meaningful ratios between arbitrary pairs of numbers. There is a rational zero point for the scale which is necessary for the ratio statements to have meaning. Example: Length in feet; duration in seconds	All statistics permitted for interval scales plus geometric mean and harmonic mean.	Same as interval plus coefficient of variation.
* The Fahrenheit scale is an example of an interval scale because each degree represents an equivalent interval and the zero point (zero degrees) is not a true zero point (i.e., there is still heat at zero degrees). With these scale properties, it is not possible to state meaningful ratios between arbitrary pairs of numbers.				
Sources: Kerlinger (1986); Leedy (1997); Tull and Hawkins (1980); Churchill (1979); Zikmund (1991)				

Table 1-2: Four Combinations of Measurement and Data

Combination	Measure Example	Rationale
Quantitative Measure Objective Data	Target Location Error	The measure is quantitative because the measurement scale is ratio (i.e., distance measured in feet). The data is collected objectively using a tape measure.
Quantitative Measure Subjective Data	Operator estimate of the probability of survival	The measure is quantitative because the measurement scale is ratio (i.e., probability ranges from 0.0 to 1.0). The data is collected subjectively by the expert responding to questions.
Qualitative Measure Objective Data	Color of munitions	The measure is qualitative because the measurement scale is nominal (i.e., blue, red, or green). The data is collected objectively through direct observation of the color or measuring the wavelength of light.
Qualitative Measure Subjective Data	Operator rating of the display	The measure is qualitative because the measurement scale is ordinal (i.e., Completely Useless, Somewhat Useless, Somewhat Useful, and Completely Useful). The data is collected subjectively by operators responding to questionnaire items.

1.3. Designing Good Questions

Regardless of whether the survey is administered by an interviewer or self-administered, good questions and proper administration are essential to collecting meaningful data. Good questions are unmistakably clear, precise, and unambiguous and ensure the recorded responses align with what the analyst is trying to measure. Questions are specifically worded to avoid creating different interpretations of what is being asked. Differences in answers should be due to differences among respondents rather than from different interpretations of the question's wording. If respondents do not have the same understanding of what a question is asking, then errors are likely. Good questions are reliable (i.e., provide consistent responses in comparable situations) and valid (i.e., answers correspond to what they are intended to measure).

There are two general types of questions that should be considered in the survey: closed and open. It is important to note that the data that can be collected is intrinsically linked to whether the question is open or closed. Closed questions provide a list of acceptable responses to the respondent. For closed questions, the analyst must develop a measurement scale to record responses. Measurement scales, usually referred to as scales, enable the analyst to measure various aspects of interest in the study.

With closed questions, respondents can answer questions more reliably since the responses are given and analysts can more reliably interpret the meaning of the answers (Fowler, 1993, p. 82).

In contrast, open questions do not provide a list of acceptable responses to the respondents. One benefit of using open questions is that the responses tend to describe more closely the real views of the respondents and can elicit unanticipated responses. Open questions are appropriate:

- When the list of possible responses is long, making it impractical to present to the respondents,

- When all reasonable responses cannot be anticipated by the interviewer or analyst.

Table 1-3 shows closed and open versions of the same question. With the closed version of the question, the responses are limited to the four that are shown. The respondent's focus is framed by providing the list of acceptable responses. With the open version of the question, the respondent may not only provide responses that are similar to those provided in the closed version of the question, but also other responses, perhaps unexpected, regarding the functions of the system.

Table 1-3: Example of Closed and Open Versions of the Same Question

Closed Version		
1. Which of the following functions can the system perform?		
Detect Threat	Yes	No
Identify Threat	Yes	No
Track Threat	Yes	No
Target Threat	Yes	No
Open Version		
2. What functions can the system perform?		

The wording of a question is critically important to ensuring the respondents' interpretations of the question are the same. The following are some recommendations for designing questions:

Use Complete Questions

Table 1-4 shows examples of both incomplete and complete versions of the same question. With incomplete questions, there is a risk that the respondents will interpret the meaning of the question differently. In question 3, age can be interpreted as being one's age as of this current time or age on one's last birthday. In question 4, one can interpret the reason to be anything and not just a medical problem or reason. The complete versions of the questions eliminate these potential interpretation problems.

Table 1-4: Examples of Incomplete and Complete Questions

Incomplete Question	*Complete Question*
3. Age?	3. What was your age on your last birthday?
4. Reason last saw doctor?	4. What was the medical problem or reason for which you most recently went to the doctor?
Source: Fowler (1993, p. 71)	

Use Specific Wording

Table 1-5 shows examples of non-specific and specific versions of the same question. In the non-specific version of question 5, the particular gun control legislation in question is not clear (i.e., banning the legal sale of certain types of guns, limiting the number and types of guns that can be owned, limiting people with certain backgrounds from owning or possessing guns). The specific version of the question fully describes the gun control legislation to ensure the respondents have a consistent interpretation.

In non-specific question 6, there are multiple appropriate responses to this question (e.g., When I was in the Army, When I was a child, When I was 4 years old, In 1985), requiring the respondent to guess what

response is required. In survey research, the question must elicit comparable answers that can be organized and analyzed. The question must specify the focus of the answer. In the specific version of the question, it is explicit that the respondent's age is the required answer.

The nonspecific version of question 7 is a "why" question that can be answered in many different ways. For instance, the respondent may remark about the capability, cost, or risk associated with Alternative C as the main factor in selecting it. The specific version of the question explicitly asks what system capabilities were considered in the selection of the alternative. In this version of the question, the respondent is prompted to address the capabilities of the system rather than the cost and risks.

There are some ambiguous words that should be avoided such as often, occasionally, usually, regularly, and frequently. These words have different meanings depending on the respondent. For example, one respondent may think regularly means three times a week, while another may think it means every day.

Table 1-5: Examples of Non-Specific and Specific Questions

Non-Specific Question	*Specific Question*
5. Do you favor or oppose gun control legislation?	5. One proposal for the control of guns is that no person who ever had been convicted of a violent crime would be allowed to purchase or own a pistol, rifle, or shotgun. Would you oppose or support legislation like that?
6. When did you have the measles?	6. How old were you when you had the measles?
7. Why did you select Alternative C?	7. What capabilities of Alternative C led you to select it?
Source: Fowler (1993, pgs. 75, 78, 79)	

Maintain Relevance

Questions in an interview or questionnaire should have a purpose that is relevant to the study. There is a natural temptation to add questions, especially "nice to know" types of questions (O'Brien and Charlton 1996, p. 89). Questions that are not relevant to the study will impose an unnecessary burden on the respondents and data analysts and should be avoided.

Check the Assumption Underlying the Question

Though a question may be clear and straightforward, the assumption underlying the question may be flawed (Leedy 1997). In the question 8 example shown in Table 1-6, the underlying assumption is that a person smokes about the same number of cigarettes each day. For some smokers, the number of cigarettes smoked will vary depending on the circumstances. For example, an impulsive smoker may smoke more cigarettes in a pressure-packed workday than at home or on holiday. Checking the assumptions underlying the questions will help ensure the questions elicit the answers the analyst is seeking. A survey pretest (discussed in Section 1.7) can help identify questions that have flawed underlying assumptions.

If a question has a flawed underlying assumption, it may be possible to revise the question to eliminate the problem. In the example shown in Table 1-6, adding a preceding question (8a) that asks whether the smoker is a consistent smoker would ensure that only consistent smokers answer question 8b.

Table 1-6: Example of a Flawed Assumption Underlying a Question

Flawed Assumption Question				
8. How many cigarettes do you smoke each day? (Check one of the following)				
1 – 5	6 – 10	11 – 15	16 – 25	More than 25
Possible Solution				
8a. Are your daily smoking habits reasonably consistent; that is, do you smoke about the same number of cigarettes each day?				
Yes		No (If "No", skip the following question)		
8b. How many cigarettes do you smoke each day? (Check one of the following)				
1 – 5	6 – 10	11 – 15	16 – 25	More than 25
Source: Leedy (1997, p. 192)				

Avoid Difficult Vocabulary

The interviewer should avoid using jargon, acronyms, or overly technical terms. Choosing words that communicate thoughts fully, clearly, and accurately through plain discourse not only helps enhance readability and comprehensibility, but also avoids ambiguity.

Avoid Exceptionally Lengthy Questions

As a general rule, the longer it takes to say something, the weaker the communication. Question length usually increases complexity and diminishes clarity, resulting in greater opportunities for misunderstanding. Single sentence questions are usually best; however, the meaning of a question or item should not be sacrificed for brevity (DeVellis 1991). Table 1-7 shows lengthy and brief versions of the same question.

Table 1-7: Example of Lengthy and Brief Versions of the Same Question

Lengthy					
What is your level of agreement with the following statement: 9. It is fair to say that one of the things I seem to have a problem with much of the time is getting my point across to other people.					
Completely Disagree	Substantially Disagree	Slightly Disagree	Slightly Agree	Substantially Agree	Completely Agree
Brief					
What is your level of agreement with the following statement: 10. I often have difficulty making a point.					
Completely Disagree	Substantially Disagree	Slightly Disagree	Slightly Agree	Substantially Agree	Completely Agree
Source: DeVellis (1991, p. 57)					

Avoid Negative and Positive Questions

Agreement bias is the tendency to agree with a question or survey item irrespective of its content (DeVellis 1991).[4] This is particularly a concern when questions are positively or negatively worded. Table 1-8 shows examples of negative and positive questions. In the negative version of question 11, respondents would tend to agree that the system has no limitations. In addition, the use of "no" in the question can be missed by the respondents. The neutral version of question not only eliminates the potential agreement bias, but is more easily understood. Similarly, in the positive version of question 12, respondents would tend to agree that the system is adequate. The neutral version of question 12 eliminates the potential agreement bias.

Table 1-8: Examples of Negative, Positive, and Neutral Questions

Negative	*Neutral*
11. Do you agree that the system has no capability limitations?	11. Rate the capability limitations of the system
Positive	*Neutral*
12. Do you agree that the system is adequate?	12. Rate the adequacy of the system

Avoid Double-Barreled Questions

Double-barreled questions pose two or more questions simultaneously. Table 1-9 shows an example of a double-barreled question. If the responsiveness and reliability of the system differ, how should the respondent answer? In this example, the problem is rectified by using two separate questions, one for responsiveness and one for reliability.

Table 1-9: An Example of a Double-Barreled Question

Double-Barreled	*Separate*
13. Rate the responsiveness and reliability of the system	13a. Rate the responsiveness of the system 13b. Rate the reliability of the system
Source: O'Brien and Charlton 1996, p. 88	

Avoid Leading and Loaded Questions

Leading questions presuppose some event or state and can bias the responses for a question (O'Brien and Charlton 1996). The leading question example in Table 1-10 presupposes that the system is unresponsive. By removing the "lack of" from the question as shown in the neutral version, the leading aspect is eliminated.

Similar to leading questions, loaded questions have emotional or sensitive content that can have carry-over effects on the entire interview or questionnaire.[5] In the loaded question example in Table 1-10, the respondent may get the impression that he or she lacks ability to perform his or her duties. In the neutral version of the question, there is no sensitive content (i.e., "your lack of ability") and the focus is directed to the system rather than the respondent.

[4] Agreement bias is also known as acquiescence or affirmation bias.

Table 1-10: An Example of Leading and Loaded Questions

Leading	*Neutral*
14. Rate the lack of responsiveness of the system	14. Rate the responsiveness of the system
Loaded	*Neutral*
15. Rate your lack of ability with respect to the duties you carried out	15. Rate your ability to perform your duties using the system
Source: O'Brien and Charlton 1996, p. 89	

1.4. Question Sequence

Opening questions that are interesting, simple to understand, and easy to answer can help establish cooperation and maintain involvement in answering questions (Zikmund 1991, p. 419). These types of questions not only help build curiosity and confidence in the respondent, but also rapport between the respondent and interviewer. Though simple and easy to answer, demographic-type questions should not be used as opening questions. Demographic-type questions tend to elicit more personal or sensitive information and should be asked later or at the end of the interview or questionnaire when a rapport between the interviewer and respondent has been established (Zikmund 1991).

The sequencing of questions and answers can create an order bias that can distort survey results (Zikmund 1991). For example, specific questions asked earlier in an interview or questionnaire can influence responses to more general questions asked later. To mitigate this bias, the analyst can use the funnel technique which entails asking more general questions to first understand the respondent's frame of reference before asking more specific questions.

Table 1-11 shows an example of how the position of questions can bias the results. In this example, the researchers found that the responses to the pollution problems were almost identical. The "air

pollution from automobile exhausts" question biased the responses to the other air pollution questions. This is an example of the anchoring effect which occurs when the first concept measured becomes the comparison point from which subsequent responses are made. To mitigate the anchoring effect, the analyst can randomize the order of the questions. In Table 1-11, for example, the order of the problem statements would vary from respondent to respondent.

Another related problem the analyst must contend with is the tendency of respondents to select the first answer listed (Zikmund 1991). Randomizing the order of responses can help mitigate this bias.

[5] Carry-over effects are produced when early questions influence or bias a respondent's answers to later questions (O-Brien and Charlton, 1996).

Table 1-11: Example of Previous Question Bias

16. Circle the number that best expresses your feelings about the severity of each environmental problem:					
Problem:	Not a Problem				Very Severe Problem
Air pollution from automobile exhausts	1	2	3	4	5
Air pollution from open burning	1	2	3	4	5
Air pollution from industrial smoke	1	2	3	4	5
Air pollution from foul smoke	1	2	3	4	5
Source: Zikmund (1991, p. 421)					

1.5. Scale Considerations

While the question prompts the response, the scale determines the form of the response. The thought and deliberation that goes into crafting good questions applies as well to selecting the appropriate scales to use. A scale must align with the wording used in the question and the intent of the measure. A question asking if something is useful, for example, should have a scale that measures usefulness in some way. If the data will be used to assess a measure (e.g., MOE, MOS, or MOP), the scale should be linked to the measure to ensure the data can be used to make an assessment as to whether the measure criteria are met or not.

When determining what response alternatives to use, the analyst should consider the following:

- Response alternatives should retain the same directional order for all questions to avoid response errors (i.e., low to high, or high to low), unless there is a belief that the order will make a difference in the responses selected.
- Balanced scales such as the Likert scale (discussed in Section 1.5.2) have an equal number of positive and negative response alternatives and tend to produce distributions that are more nearly normal (O'Brien and Charlton, 1996, p. 84).
- Although greater discriminability can be obtained by adding more response alternatives, more than seven response alternatives increases response variability and lowers overall reliability (O'Brien and Charlton, 1996, p. 87).

The response alternatives (also known as scale descriptors) used in scales are chosen for consistency, discriminability, and comprehensibility. Response alternatives with these attributes can help avoid nonresponse and response bias. Appendix E shows examples of five, six, and seven point scales with these attributes.

1.5.1. Nominal and Ordinal Scales

Closed questions typically have nominal or ordinal based scales to measure a response. As shown in Table 1-12, the nominal scale uses categories (e.g., yes or no) that have no rank order relationship. In contrast, the ordinal scale uses a rank order relationship. Responses to questions are normally treated as nominal or ordinal data based on the scale used. It is important that the analyst know the type of data and the appropriate statistics that can be used (see Section 1.2 (Types of Measurement and Data) for more information).

Table 1-12: Nominal and Ordinal Based Scales

Nominal Scale	
The system will enable the nuclear enterprise to accomplish its mission.	
Yes	No

Ordinal Scale					
Rate the adequacy of the system in supporting the nuclear enterprise mission.					
Totally Inadequate	Very Inadequate	Somewhat Inadequate	Somewhat Adequate	Very Adequate	Totally Adequate

1.5.2. Likert Scale

The Likert scale, developed by sociologist Rensis Likert to measure psychological attitudes in a scientific way, is an ordinal based scale that is commonly used in studies to measure the level of agreement or disagreement. As shown in Table 1-13, the scale is bivalent (two-directional) and balanced (i.e., equal number of positive and negative response alternatives) with a neutral middle. The scale has verbal labels that connote evenly spaced graduations of the response alternatives. Five-point response alternative scales are often used, though seven and nine point scales can be used as well.

Table 1-13: Example of a Likert Scale

A two-level maintenance concept can be used to maintain this system.				
Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree

1.5.3. Neutral Midpoint

Another aspect the interviewer must consider is whether to use a neutral midpoint in a scale such as the example shown in Table 1-13 above. Scales without a neutral midpoint force respondents to select a response that departs from true neutrality which can occasionally result in nonresponse. The drawbacks of forcing respondents to make a choice must be carefully weighed against the benefits of obtaining non-neutral responses.

1.5.4. Not Applicable or Don't Know as a Selection

The selection of respondents to participate in a survey not only requires careful planning and preparation, but also a thorough understanding of the respondents' qualifications to answer the survey questions. Despite the best efforts of the interviewer, there may be cases when respondents are asked questions concerning things about which they do not know. One approach to deal with such a possibility is to include "Not Applicable" or "Don't Know" as a selection separate from the response alternatives. A "Not Applicable" or "Don't Know" selection indicates the respondent did not have adequate knowledge or experience on which to base an answer. These selections are for administrative purposes and are separate from the responses of interest.

There are two other reasons for including "Not Applicable" or "Don't Know" as a selection. First, it will allow the analyst to better ascertain whether item nonresponse was intentional or unintentional since the likelihood that respondents will intentionally not answer the question is low. Second, a "Not

Applicable" or "Don't Know" selection helps prevent respondents who do not have adequate knowledge or experience on which to base an answer from arbitrarily selecting a response. Table 1-14 shows an example of an item with "Don't Know" included as a selection.

When qualified respondents do select a "Not Applicable" or "Don't Know" response, the analyst must investigate the reason for the selection to ensure it was not accidental and that the respondent actually had no basis for an opinion. In the report, the analyst should clearly identify these occurrences, describe the reasons for them, provide justification for their removal from the sample population, and clearly document the actual sample size. Because neutral responses like "Neither Agree nor Disagree" or "Borderline" are not particularly informative, the analyst should also investigate and document the reasons for these response selections. Finally, the analyst should always investigate and document any response anomalies such as outliers (in either direction) and bimodal distributions. With graphical and numerical representations of the data, the analyst can discover patterns and anomalies and identify potential causes.[6] The information helps facilitate understanding and interpretation of the data, enabling the analyst to describe and present results in more meaningful ways.

Table 1-14: Item with "Don't Know" Included as a Selection

How important is the airborne radiation survey reconnaissance map for the ground planning mission?					Don't Know
Not Important at All	Not so Important	Neutral	Fairly Important	Very Important	
Comments:					

1.6. Designing the Survey

Whether the survey is interview-administered or self-administered, there are several characteristics of a well-designed questionnaire. Though the questions may be worded properly and have the appropriate format and response scale, the questions should flow from the most general to the most specific, or from the most frequent/common events to the rare or unusual (O'Brien and Charlton 1996, p. 89). This is done to minimize carry-over effects, instances where the responses to earlier questions bias responses later in the questionnaire.

A well-designed questionnaire is brief and to-the-point (O'Brien and Charlton 1996). With overly long questionnaires, there is a risk of respondents choosing neutral or extreme responses out of expediency or fatigue. If the questionnaire takes more than 15 minutes, O'Brien and Charlton (1996) recommend that the analyst consider dividing the questions into two or more questionnaires that can be administered at different times during the study.

The remaining discussion in this section describes guidelines for designing the interview-administered survey and self-administered survey.

---

[6] Examples of graphical representations include bar charts, histograms, and box and whisker plots. Numerical representations include measures of central tendency such as the mean, median, and mode and measures of variability such as range, variance, and standard deviation.

1.6.1. Interview-Administered Survey

In the interview-administered survey, an interviewer reads the questionnaire instructions and questions to the respondent and records the responses. Fowler (1993, p. 99) offers the following guidelines for designing a survey that will not only make the tasks of the interviewer and respondent as easy as possible, but also help ensure the information elicited from the respondents is what the interviewer is seeking:

- Adopt a convention that differentiates between the words that the interviewer reads to respondents and words that are instructions. A common convention is to use uppercase letters for the instructions and lowercase letters for words to be read aloud.
- Establish a clear convention for handling instructions to skip questions that do not apply to a particular respondent. The instructions should be keyed to a particular response and tell the interviewer where to go to ask the next questions. Visual cues such as boxes and arrows are probably the most self-explanatory. Such visual cues, however, require some formatting at the typing and printing stages of the questionnaire. Another option is to use clearly written "skip" instructions. Whatever approach is used, it is worthwhile to be completely consistent so that the interviewer does not have to spend time thinking about which questions to ask.
- Put optional wording in parentheses. Conventions such as "(his/her)" or "(husband/wife)" are easy for interviewers to handle smoothly if they are alerted by the parentheses. A similar convention uses all caps (e.g., NAME) when the interviewer must supply a word that is not provided in the question itself.
- Check to make sure that all the words that an interviewer has to say are, in fact, written. This includes not only the phrasing of questions but transitions, instructions to questions, needed definitions, and explanations.

An additional interview technique to consider is the pitcher/catcher routine whereby one interviewer pitches a question while a second interviewer records the response. This technique allows the pitcher to focus on the response and generate follow-up questions if necessary. Additionally, the chances of missing a response or parts of a response are greatly reduced by having one interviewer devoted to recording the response.

1.6.2. Self-Administered Survey

In the self-administered survey, the respondent reads the questions in the questionnaire and provides responses, hence the name "self-administered". Similar to interview-administered surveys, self-administered surveys should have a questionnaire that is easy to use, although the format and design of the self-administered surveys is typically more important. Self-administered surveys should have a cover sheet that includes the title of the questionnaire, purpose of the questionnaire, information regarding the use of the data, tracking information (e.g., date, time, location), and if needed, clear and brief instructions and an assurance of confidentiality.

For self-administered surveys, Fowler (1993, p. 100) offers the following recommendations:

- A self-administered questionnaire should be self-explanatory. Written instructions should not be necessary because they will not be read consistently.
- Self-administered questionnaires should be restricted to closed questions. Checking a box or circling a number should be the only tasks required. When respondents are asked to answer in their own words, the answers are usually incomplete, vague, and difficult to code, and therefore they are of only limited value as measurements.[7]

- The question forms in a self-administered questionnaire should be few in number. The more the questionnaire can be set up so that the respondent has the same kinds of tasks and questions to answer, the less likely it is that respondents will become confused; also, the easier the task will be for respondents.
- A questionnaire should be typed and laid out in a way that is clear and uncluttered. Photo-reduction (or other strategies for putting many questions on a page) actually reduces the response rate compared to when the same number of questions are spaced attractively over more pages.
- Question skip patterns (i.e.., how respondents skip questions that do not apply to them) should be kept to a minimum. If some respondents must skip some questions, arrows and boxes that communicate the skips without verbal instructions are best.

1.7. Pretesting the Survey

There are several approaches to pretest a survey. For interview-administered surveys, one approach entails identifying and selecting qualified individuals (i.e., subject matter experts or others familiar with the area of interest in the study) to serve as respondents for pre-testing the questions. The interviewer conducts trial interviews to observe whether the respondents have difficulty answering the questions and determine whether the responses align with expectations. For example, a respondent's request for additional information may indicate a problem with the question sequence or the question itself. The interviewer can test alternative wordings, order and number of questions, question format (e.g., open versus closed), and sequence to determine which approach is best suited for the respondents.

A similar approach can be used to pretest self-administered questionnaires. The interviewer first identifies and selects qualified individuals to serve as respondents. Once the respondents have been selected, the interviewer directs the respondents to fill out the questionnaire. Finally, the interviewer leads a discussion to determine whether there are any problems regarding the clarity and completeness of the instructions, expected responses (i.e., whether the responses align with expectations), and question wording, format, and order.

As part of pre-testing the survey, the interviewer determines whether the length of the interview or questionnaire is appropriate. The criteria for determining survey length include cost, the effect on response rate, and limits of the respondent's ability and willingness to answer questions (Fowler 1993, p. 103).

1.8. Sampling

There are two main types of sampling: nonprobability and probability. This section describes each type and provides recommendations for determining sample size.

1.8.1. Nonprobability Sampling

Unlike probability sampling, nonprobability sampling is a procedure for selecting the sample elements that is not based on chance (i.e., the selection is not random). A special form of nonprobability sampling commonly used in operational capability requirements studies is judgment or purposive sampling.[8]

[7] Some experts in survey design recommend providing a comment section with each closed question to provide an opportunity for the respondent to explain the rationale behind a response.

In purposive sampling, the analyst uses judgment in selecting the sample elements (i.e., those who the analyst intends to interview or will fill out a questionnaire). The analyst purposively selects the sample elements because they will serve a specific purpose and have some appropriate characteristic that is required. There are a variety of reasons the analyst can use to select the sample elements. For example, there may be a limited number of individuals who have the expertise in the area being studied, or the interest of the study is on a small group or in a specific field. Alternatively, the analyst may be attempting to draw a representative sample of a population, but uses judgment rather than a probabilistic approach in the selection procedure.

Though there are many advantages of nonprobability sampling, all nonprobability sampling techniques are subject to selection error (Tull and Hawkins 1980, p. 389). The level of selection error is dependent on the degree of expertise of the individual making the selection decision. As the sample size increases, judgment becomes less trustworthy compared to random sampling procedures used in probability sampling. In addition, given that the elements are not selected probabilistically, there are no appropriate statistical techniques for measuring sampling error (i.e., the degree to which a parameter (e.g., median, mean) of a sample represents the parameter of a population) and projecting the data beyond the sample (Zikmund 1991, p. 462).

1.8.2. Determining the Sample Size for Nonprobability Sampling

There is little theoretical basis for determining the sample size for nonprobability sampling. With the sample elements being selected based on judgment, some scholars think sample size may also be determined on the basis of judgment, though no heuristics or criteria have been prescribed (Zikmund 1991, p. 510). Others recommend using sample sizes that are similar to sample sizes used in previous studies.

Another approach known as the Bayesian method for determining sample size entails choosing a sample size based on the expected net gain of sampling. The expected net gain of sampling is the difference between the expected value of information and the cost of sampling. The objective is to choose the sample size that produces the greatest positive difference. Though it may be possible to calculate the cost of sampling, calculating the expected value of information is much more difficult and may require resources beyond those of a typical capability requirements study.

A more practical approach is to calculate the sample size as if it were a probability sample. The procedure described in Section 1.8.4 (Determining the Sample Size for Probability Sampling) can be used to determine the sample size for nonprobability sampling.

1.8.3. Probability Sampling

In some situations, the analyst may not only want the sample elements to represent a population of interest, but also a selection procedure based on chance (i.e., the selection is random). For example, an analyst may want to first gather information about a particular population (e.g., F-16 aircraft pilots), and then select individuals from the population to participate in an expert elicitation panel to gather more specific information. In this example, probability sampling could be used to gather the information about

[8] Another type of nonprobability sampling is convenience sampling (selecting individuals that are most conveniently available). Convenience sampling is also known as accidental or haphazard sampling. Subcategories of purposive sampling include expert sampling (selecting individuals with known or demonstrable experience and expertise in some area) and quota sampling (selecting individuals non-randomly according to some fixed quota where the quota may be proportional or non-proportional).

the F-16 pilot population. In probability sampling, every individual in the population of interest has a chance to be selected as a sample element. The procedure for selecting the sample elements is random and is the basis for all probability sampling techniques.

There are several different types of probability sampling techniques. Simple random samplings, simple and proportional stratified sampling, and cluster sampling, are some examples of probability sampling techniques. As shown in Table 1-15, selecting the appropriate sampling technique requires an understanding of the population.

Simple random sampling is the least sophisticated of all the probability sampling techniques. The sample is derived by means of a simple randomization process. Random number tables or computer- generated numbers are used for the sample selection. More information about simple random sampling and the other types of probability sampling techniques can be found in the references listed in Appendix C.

1.8.4. Determining the Sample Size for Probability Sampling

With probability sampling, whatever statistics are calculated from the sample (e.g., mean, median, percentage), a population value is being estimated. The relationship between sample size and sampling error (i.e., the degree to which a parameter of a sample represents the parameter of a population) can be generally expressed as: the smaller the sample, the larger the error; and the larger the sample, the smaller the error (Kerlinger 1986). As a general rule, the analyst should use as large a sample as possible.

There are statistical equations that can be used to estimate sample size. The choice of which equation to use is dependent on whether the data being collected is attribute or variable data. Table 1-16 shows examples of survey questions designed to elicit attribute data and variable data. Attribute data is associated with the proportion or percentage of the population that has the same characteristic, feels a certain way, or views something the same way. As shown in the table, the attribute question is

designed to elicit a maintainer's opinion about how easy or difficult it would be to maintain the new electronic warfare system. In this example, the analyst is interested in knowing what percentage of the respondents think the system would be somewhat easy or extremely easy to maintain.

Table 1-15: Sampling Techniques Appropriate for Each Population Type

Population Characteristic	Example	Appropriate Sampling Technique
Population is generally a homogeneous group.	A quantity of flower seeds of a particular variety from which random samples are selected for testing as to their germination quality.	Simple Random Sampling
Population consists of definite strata, each of which is distinctly different, but the units within the stratum are as homogeneous as possible. Each stratum is essentially equal in size.	A particular town whose total population consists of three types (strata) of citizens: European-American type; African-American type; and Mexican-Indian type.	Simple Stratified Sampling
Population consists of definite strata with differing characteristics, and each stratum has a proportionate ratio in terms of numbers of members to every other strata.	A community in which the total population consists of individuals whose religious affiliations are found to be Catholic, 25%; Protestant, 50%, Jewish, 15%; nonaffiliated, 10%.	Proportional Stratified Sampling
Population consists of clusters whose cluster characteristics are similar yet whose unit characteristics are as heterogeneous as possible.	A survey of the nation's 20 leading air terminals by soliciting reactions from travelers who use them. (All air terminals are similar in atmosphere, purpose, design, and so forth, yet the passengers who use them differ widely in individual characteristics: age, gender, national origin, philosophies, beliefs, and socioeconomic status)	Cluster Sampling
Source: Leedy 1997, p. 214		

Table 1-16: Examples of Questions Designed to Elicit Attribute Data and Variable Data

Attribute Data Question				
How easy or difficult would it be to maintain this new electronic warfare system?				
Extremely Difficult	Somewhat Difficult	Borderline	Somewhat Easy	Extremely Easy
Variable Data Question				
How long will it take in hours to perform routine scheduled maintenance on the new electronic warfare system?				
Write the number of hours here:				

In contrast, variable data is associated with an amount of something expressed as a number or value. The

variable question example in the table is designed to elicit a value from the respondents. The analyst is interested in determining the average number of hours from the estimates provided by the respondents.

In capability requirements studies, the analyst is typically dealing with attribute data rather than variable data such as means and standard errors of a mean. For attribute data, the analyst would use equations designed to estimate sample size involving proportions. The remaining discussion in this section describes how the analyst would use these equations to determine the sample. Appendix F provides the equations for estimating sample size involving means. Though different equations are used, the approach is similar to what is described in this section.

According to Emory (1985, p. 287), the single most important factor in determining the size of a sample needed for estimating a parameter of a population is the magnitude of the population variance (commonly expressed as the standard deviation), not the size of the population. The greater the dispersion or variance of the population, the larger the sample must be to provide an estimate with a given level of accuracy and precision. For example, if a population held the same view about something (i.e., no variance), then only a sample of one would be needed to know the view held by the population. If there are many possible views, then a larger sample is needed to collect them and estimate their frequencies.

For the statistical equations that are used to estimate sample size, an estimate of the population variance is needed (Zikmund 1991, p. 506).[9] Information from prior studies or a pilot study is commonly used to estimate the variance. In capability requirements studies, it is likely that such information is not available or time and resource constraints make it impossible to collect the information through a pilot study. With attribute data, it is still possible to calculate a sample size.

The variance is measured in terms of *p*, the proportion of the population that has the given attribute. With *n* = size of the sample, the variance of the proportion is as follows:

$$\sigma_p^2 = \frac{p(1-p)}{n}$$

The standard error of the proportion is as follows:

$$\sigma_\rho = \sqrt{\frac{p(1-\rho)}{n}}$$

Solving for *n*, the equation becomes:

$$n = \frac{p(1-p)}{\sigma_\rho^2}$$

Before the sample size can be estimated statistically, the analyst must determine (1) the magnitude of acceptable sampling error (precision); and (2) the confidence level. These specifications typically involve judgment and are based on how the data will be used. Questions such as "How much error in the estimate is acceptable?" and "How confident do you want to be that the error really isn't greater than that?" must be answered (Tull and Hawkins 1980, p. 413).

[9] See Zikmund (1991), Emory (1985), Tull and Hawkins (1980), and O'Brien and Charlton (1996) for variations of statistical equations for estimating the sample size for variable data (e.g., mean and standard errors of the mean) and attribute data.

As noted earlier, sampling error is a measure of precision and is the degree to which a parameter (e.g., median, mean) of a sample represents the parameter of a population. Sampling error is also referred to as the confidence interval. The sampling error or confidence interval is a specified range of numbers within which a population parameter should lie. For example, a sampling error of plus or minus 10% means the sample parameter is within 10% of the population parameter.

Closely associated with the sampling error (or confidence interval), the confidence level is distinct and calculated differently. Expressed as a percentage or decimal value, the confidence level tells how confident an analyst can be about being correct (Zikmund 1991). For example, a confidence level of .95 means there is a .95 probability that the true population parameter is correctly estimated. Stated another way, there is a .05 probability that the true population parameter is incorrectly estimated.

Using judgment, the analyst selects the appropriate items or questions to be used for sample size calculations (Zikmund 1991, p. 510). In most studies, the desired degree of precision may be different across the items or questions selected by the analyst. The convention is to select the item or question that will produce the largest sample size and use it to determine the ultimate sample size. The analyst should consider the cost of data collection and exercise judgment regarding the importance of such information.

To facilitate an understanding of determining a sample size involving proportions, an example for simple random sampling[10] is described below. Appendix F provides the equation used to calculate the sample size involving means.

EXAMPLE

Using a simple random sampling technique, the analyst wants to determine the size of the sample to estimate the true proportion in the population within plus or minus 10 percent. In addition, the analyst wants to be 95 percent confident that the population proportion is within plus or minus 10 percent of the sample proportion. Given that there is no estimate of the population variance, the analyst follows the established convention of using the largest variance. Recall that with attribute data, that variance is measured in terms of p, the proportion of the population that has the given attribute. In the standard error of a proportion equation shown below, the standard error is largest when $p = 0.5$. For example, if $p = 0.5$, then the product of p (0.5) and $1 - p$ (0.5) is 0.25. If p or $1 - p$ is greater than 0.5, then the product will always be smaller than 0.25. If $p = 0.6$, for instance, then $p(1 - p) = 0.6 \times 0.4 = 0.24$. With this feature of the equation, the convention is to assume the largest variance by selecting $p = 0.5$ when an estimate of the variance is not known. Note that the cost of p being unknown is an increase in the sample size.

$$\sigma_\rho = \sqrt{\frac{p(1-\rho)}{n}}$$

[10] Though the principles of determining the sample size for simple random sampling are applicable to all methods of probability sampling, see Tull and Hawkins (1980, p. 422 and Appendix B) for more information about determining sample size for non-simple random sampling techniques.

[11] In a normal curve, this means that an incorrect estimate must fall in each tail of the normal curve where the proportion of the area is 0.025. With two tails in the normal curve, each with a proportion of area of 0.025, the total proportion of area becomes 0.05.

Given that *n* is what the analyst is trying to calculate in the standard error equation above, the next step is to calculate the standard error (σ_p). As noted earlier, the analyst wants to determine the size of the sample to estimate the true proportion in the population within plus or minus 10 percent. This means the confidence interval will be ± 0.1. Recall that the analyst wants only a .05 probability that the true population proportion is incorrectly estimated.[11] Hence, the confidence interval must encompass a dispersion of ± 1.96 standard errors of the proportion. The 1.96 value is the z-value that corresponds to the 0.025 proportion of the area of the normal curve. One standard error is determined by dividing the confidence interval by the standard errors (0.1/1.96 = 0.051).

With the standard error calculated, the analyst can now calculate the sample size *n* as shown below:

$$n = \frac{p(1-p)}{\sigma_p^2}$$

$$n = \frac{.5 \times .5}{.051^2}$$

$$n = 96$$

Rather than calculating the sample size for each survey, the analyst can use a table such as the one shown in Table 1-17. There are several aspects of the table that are worth noting. First, the sampling error or confidence interval is greatest when the proportion is 50/50 and decreases as the proportion approaches 5/95. Second, as the sample size increases, the sampling error decreases. It is important to note that sampling error steadily decreases as the sample size increases from 35 to 200, after which point the reductions in sampling error become much less. Adding sample elements to a sample reduces sampling error more significantly for small samples than large samples. For sample size from 35 to 75 decreases the confidence interval by 2 (from 7 to 5) for the 5/95 proportion, whereas increasing the sample from 300 to 350 would not yield a noticeable decrease in sampling error.

Fowler (1993, p.34) points out that it is unusual for an analyst to be able to specify a desired level of precision in more than a general way. It is only the exception, rather than the common situation, when a specific level of precision can be specified in advance. This means it would be unusual for the analyst to use the equations described above to determine the sample size. Given this, Fowler (1993) highlights the value of the table in determining the sample size. He recommends that the analyst start at the low end of the sample size continuum shown in the table rather than the high end. With the level of precision associated with each sample size, the analyst asks the question “Are 35 observations adequate?” If not, the analyst asks the same question for a larger sample size and so on. As noted earlier, sampling error steadily decreases as the sample size increases from 35 to 200, after which point the reductions in sampling error become much less.

Table 1-17: Confidence Ranges for Variability Attributes to Sampling

Sample Size	Proportion of Sample with Characteristic[1,2]				
	5/95	10/90	20/80	30/70	50/50
35	±7%	±10%	±14%	±15%	±17%
50	±6%	±8%	±11%	±13%	±14%
75	±5%	±7%	±9%	±11%	±12%
100	±4%	±6%	±8%	±9%	±10%
200	±3%	±4%	±6%	±6%	±7%
300	±3%	±3%	±5%	±5%	±6%
500	±2%	±3%	±4%	±4%	±4%
1,000	±1%	±2%	±3%	±3%	±3%
1,500	±1%	±2%	±2%	±2%	±2%

[1] Chances are 95 in 100 that the real population figure lies in the range defined by the ± number indicated in the table, given the proportion of sample reporting the characteristic and the number of sample cases on which the proportion is based.
[2] This table describes the variability attributable to sampling. Errors resulting from nonresponse or reporting errors are not reflected in this table. In addition, a simple random sampling is assumed. Estimates may be subject to more variability than this table indicates because of the sample design or the influence of interviewers on the answers they obtained; stratification might reduce the sampling errors below those indicated here.

Source: Fowler (1993, p. 31)

1.9. Survey Question Validity and Reliability

The approach to designing questions and surveys described in this handbook is in accordance with widely accepted practices in survey research. Practices such as the wording and length of questions, pretesting, and survey design and administration are intended to establish question reliability (provide consistent measures in comparable situations) and validity (answers correspond to what they are intended to measure).

There are statistical techniques to measure question reliability and validity, but using such techniques would likely exceed the resources of a typical capability requirements study and are beyond the scope of this handbook. For more information regarding statistical techniques for assessing validity and reliability, please see DeVellis (1991), Scale Development: Theory and Applications.

1.10. Administering the Survey

Given that the approach used to administer the survey can influence the quality of the data collected, it is important that the interviewer use proper procedures in administering the survey. This section describes some guidelines for administering surveys.

1.10.1. Interview-Administered Surveys

According to Fowler (1993), the interviewer has three primary roles to play in the collection of interview-administered survey data:

- Locate and enlist the cooperation of selected respondents,

- Train and motivate respondents to do a good job of being a respondent,
- Ask questions, record answers, and probe incomplete answers to ensure that the answers meet the question objectives.

To enlist cooperation, the interviewer must first locate potential respondents and make contact. Though some individuals may readily agree to being interviewed, others may be reluctant to participate. Interviewers who are good at enlisting cooperation typically exhibit a confident assertiveness and personable style (Fowler 1993). They present the study in a way that the individual will want to participate. The tone and content of the conversation leaves no doubt that an interview will be conducted. The interviewer engages the individual personally by tailoring the interaction to be responsive to the individual's needs, concerns, and situation. It is also important that the interviewer is available to conduct an interview when a respondent wants to be interviewed.

Most respondents have little understanding of what they are expected to do and how they should perform (Fowler 1993). A critical part of the interviewer's job is to train and motivate the respondent's behavior. Studies have shown how the encouragement provided by the interviewer affects what respondents do and how well they report. For example, an interviewer who reads a question quickly encourages the respondent to answer the question quickly, but not necessarily accurately. In contrast, an interviewer who reads a question slowly and deliberately encourages the respondent to take the time to thoughtfully and accurately answer the question.

To ensure that differences in responses are attributed to differences in the views and experiences of the respondents rather than to differences in how the survey is administered, there are aspects of interviewer behavior that must be standardized. Fowler (1993, pgs. 107-108) offers the following:

Presenting the Study. Respondents should have a common understanding of the purposes of the study. This sense of purpose may have a bearing on the way they answer questions. Assumptions about such things as confidentiality, the voluntary nature of a project, and who will use the results also can have some effect on answers. A good interviewer will give all respondents a similar orientation to the study so that the context of the interview is as consistent as possible.

- Asking the Questions. Survey questions are supposed to be asked exactly the way they are written, with no variation or wording changes. Even small changes in the way questions are worded have been shown, in some instances, to have significant effects on the way questions are answered.
- Probing. If a respondent does not answer a question fully, the interviewer must ask some kind of follow-up question to elicit a more complete answer; this is called probing. Interviewers are supposed to probe incomplete answers in nondirective ways—ways that do not push the respondent and increase the likelihood of any one answer over another. A short list of standard probes, including repeating the question, asking "Anything else?", "Tell me more", and "How do you mean that?" will handle most situations if the survey is designed well.
- Recording the Answers. The recording of answers should be standardized so that no interviewer-induced variation occurs at that stage. When an open-ended question is asked, interviewers are expected to record answers verbatim; that is, exactly in the words that the respondent uses, without paraphrasing, summarizing, or leaving anything out. In closed- response questions, when respondents are given a choice of answers, interviewers are required only to record an answer when the respondent actually chooses one. There is potential for inconsistency if interviewers code respondent words into categories that the respondent did not choose.

- Interpersonal Relations. Manage the interpersonal aspects of an interview in a standardized way. Inevitably, an interviewer brings some obvious demographic characteristics into an interview, such as gender, age, and education. By emphasizing the professional aspects of the interaction and focusing on the task, the personal side of the relationship can be minimized. Interviewers generally are instructed not to tell stories about themselves or to express any views or opinions related to the subject matter of the interview. Interviewers are not to communicate any judgments on answers that respondents give. In short, behaviors that communicate the personal, idiosyncratic characteristics of the interviewer are to be avoided because they will vary across interviewers if more than one interviewer is used. Professional behavior helps to standardize the relationship across interviewers and respondents. There is no evidence that having a friendly interpersonal style improves the accuracy of reporting; it probably tends to have a negative effect.

1.10.2. Self-Administered Surveys

Proper administration of self-administered surveys includes providing the questionnaire to respondents, ensuring the respondents understand all items, conducting a quality control check of the responses, and actively investigating reasons for certain responses (generally those that are ambiguous or unexpected). For capability requirements studies, questionnaires should never be simply handed to respondents who are then asked to "fill them out and return them whenever you can." A much more effective approach is to schedule a specific time and place for the respondents to gather and complete the questionnaires. The interviewer remains with the group to field questions and clarify items that may be confusing or are being misinterpreted. As questionnaires are returned, the interviewer should carefully examine them to ensure:

- A response (or "Not Applicable" or "Don't Know" selection) has been selected for all items,
- The respondent viewed the scale directions correctly (an indication that the scale direction may have been viewed incorrectly is when responses from a respondent are mostly opposite to those of other respondents), and
- Responses to open-ended questions and any other comments can be read and understood.

If there are issues with any questionnaire responses, the interviewer should review and resolve them with the respondent immediately. It is never good practice to put off addressing questionnaire problems to a later date as memories fade and people may become unreachable as they move on to other activities.

1.11. Data Analysis

As previously shown in Table 1-1, there are various descriptive and inferential statistics that can be used depending on the type of data (i.e., nominal and ordinal). To use these statistics, the analyst must first convert the data into a numerical form. To do this, the analyst assigns numerical values to the descriptors used in the measurement scale. Table 1-18 shows an example of a numerical value assignment for a five-point Likert scale. Using the scale and associated numerical values, the analyst scores the numerical value that corresponds to the respondent's answer. For example, an answer of "Strongly Agree" would be scored as a "5". An answer of "Somewhat Agree" would be scored as a "4", and so on.

Table 1-18: Assigning Numerical Values to the Scale Descriptors

Question:	A two-level maintenance concept can be used to maintain this system.				
Scale Descriptor:	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Assigned Value:	1	2	3	4	5

There are several graphical representations such as the bar chart, histogram, and box and whisker plot that the analyst can use that will help facilitate an understanding of the data.[12] Figures 1-1 and 1-2 show examples of the bar chart and box and whisker plot. In these examples, subject matter experts were asked to rate how easy or difficult it will be to maintain, transport, and deploy a specific system. The scale used with each question was a five-point Likert scale ranging from "Extremely Difficult" to "Extremely Easy". As shown in Figure 1-1, the bar chart is useful for showing the response frequencies of each rating and overall pattern of the data. These representations can reveal whether there are outlier responses and the nature of the distribution (e.g., bimodal). In this example, the single "Extremely Easy" response to the transportability question is a potential outlier.

The box and whisker plot is one way to illustrate both variability and central tendency of the data in a single format (see Figure 1.2). The top and bottom of each box represents the third quartile (Q3) and first quartile (Q1) respectively. The horizontal red line inside the box represents the median value. The whiskers can represent several possible values such the minimum and maximum values, the 2nd and 98th percentiles, and multiples of the interquartile range values.[13] In this example, the whiskers represent the minimum and maximum values of the data. As shown in the figure, the variability of the deployability responses (as indicated by the size of the box) is much less compared to the variability of the maintainability and transportability responses.

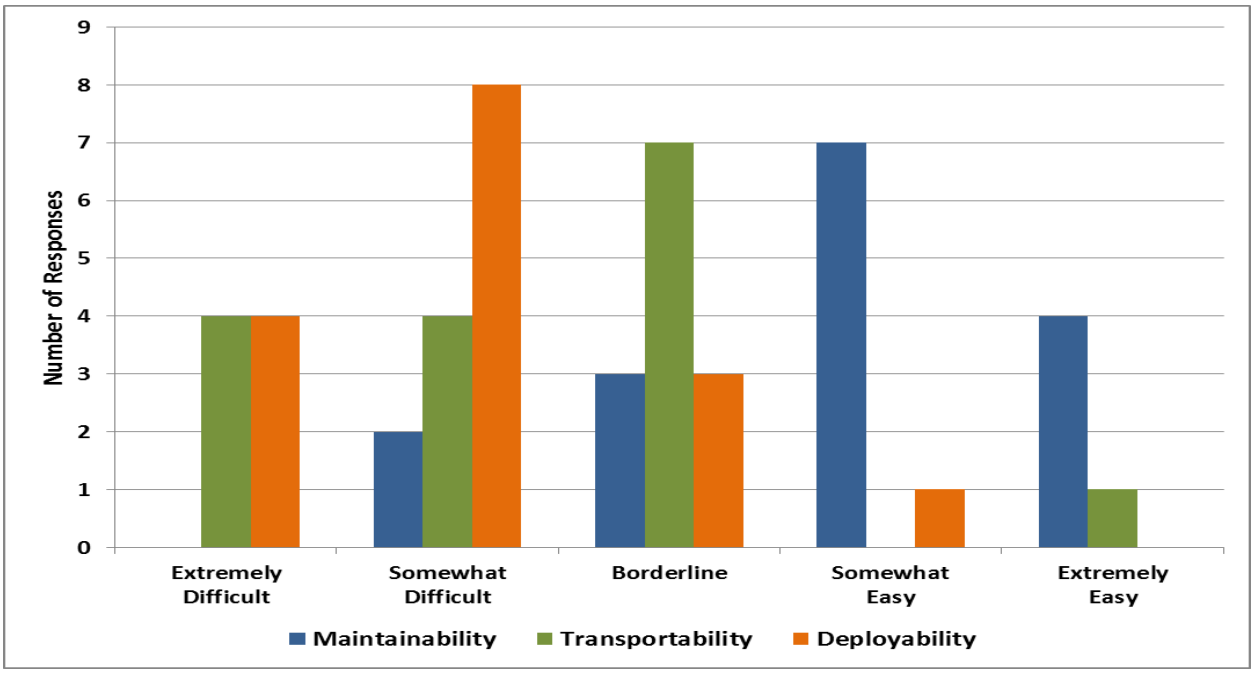


Figure 1-1: Bar Chart Example

[12] There is a difference between the bar chart and the histogram. The bar chart is used to present categorical-type data such as nominal and ordinal levels of measurement. The histogram is used to present continuous data such as that associated with interval and ratio levels of measurement.

[13] The interquartile range (IQR) is the difference between Q3 and Q1. The interquartile range is often used to find outliers in data. Outliers are observations that fall below Q1 - 1.5(IQR) or above Q3 + 1.5(IQR). In a box and whisker plot, the highest and lowest values are drawn as the bar of the whiskers, and the outliers as individual points.

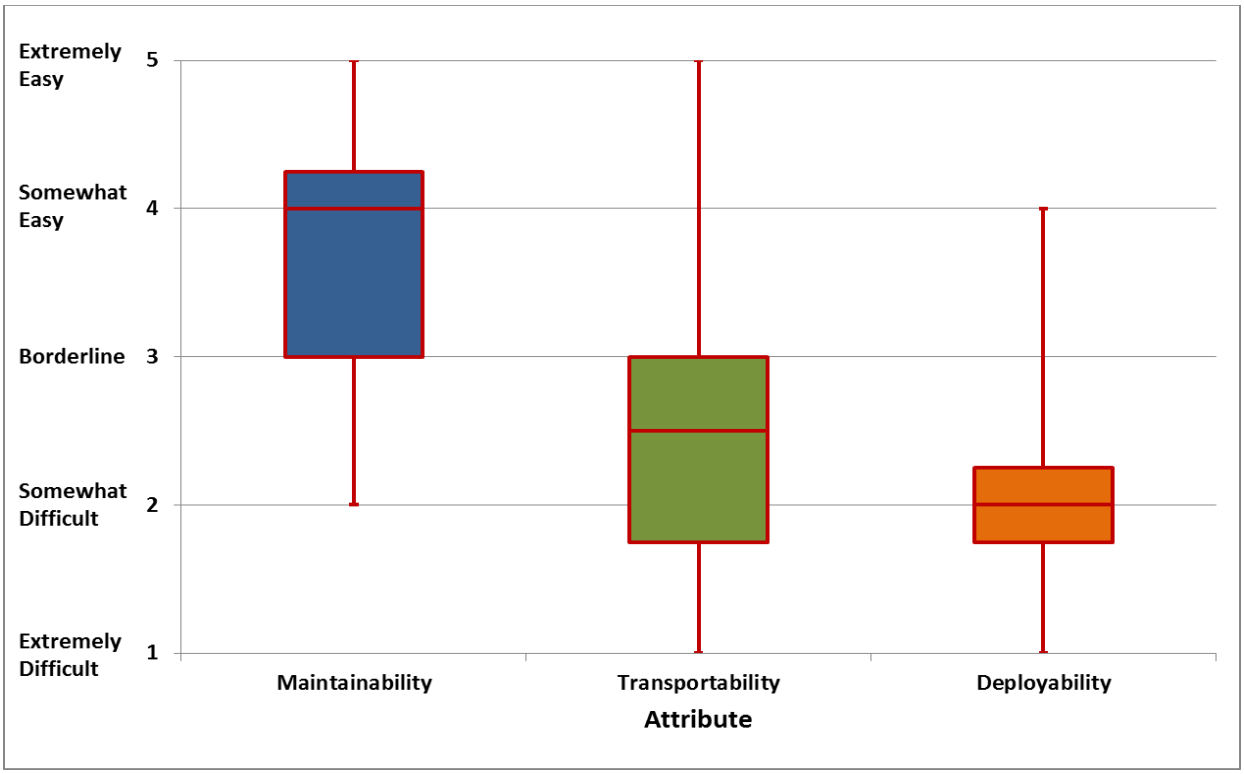


Figure 1-2: Box and Whisker Plot Example

1.11.1. Central Tendency and Dispersion Measure Criteria

The criteria used for a measure may require calculating the central tendency, dispersion, or both. Central tendency statistics such as the median or mode are used to identify the direction (positive or negative) and magnitude of the ratings. Dispersion statistics such as the variance or ratings corresponding to a percentile capture the level of agreement in the ratings.

Table 1-19 shows an example of a measure where the criteria requires the computation of a central tendency statistic and a dispersion statistic. The measure criteria states that the median rating must be greater than or equal to "Somewhat Agree" and the 80th percentile must be "Somewhat Agree" or better. In this example, the central tendency and dispersion criteria are met, indicating that not only a majority of the responses are favorable (86% in this example), but there is substantial agreement among the respondents.

Table 1-19: Examples of Central Tendency and Dispersion Measure Criteria

Measure Description

Task	Attribute	Measure	Metric	Criteria (Threshold)	Data
Maintain and Sustain System	Maintainability	Logistician Rating of Maintainability (MOS)	Median	≥ Somewhat Agree	Logistician responses to questionnaire item with a 5-point Likert scale (see below)
			Percentile	≥ 80th Percentile must be Somewhat	

Questionnaire Item and Responses

Question:	A two-level maintenance concept can be used to maintain this system.				
Scale Descriptor:	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Assigned Value:	1	2	3	4	5
Number of Responses:	0	0	2	8	4
Results Median: 4 (Somewhat Agree) Percentile: 86th Percentile is Somewhat Agree or Strongly Agree					

1.11.2. Ordinal Data Analysis

One pitfall to avoid is the inappropriate analysis of ordinal data. Recall that ordinal data is one of the types of data that is usually collected through surveys.[14] In the example shown in Table 1-20, data for the measure was collected through a self-administered survey and exhibits the properties of ordinal data (i.e., data are assigned numbers such that the order of the numbers reflects an order relation based on the attribute). The analyst, however, incorrectly selected the mean as the metric. Despite six of the seven respondents agreeing that a two-level maintenance concept can be used for the system, the measure is rated "did not meet criteria" because 3.8 is not greater than or equal to the threshold value of 4 (Somewhat Agree).

Table 1-20: Inappropriate Analysis of Ordinal Data Example

Measure Description

Task	Attribute	Measure	Metric	Criteria (Threshold)	Data
Maintain and Sustain System	Maintainability	Logistician rating of maintainability (MOS)	Mean	≥ Somewhat Agree	Logistician responses to questionnaire item with 5-point Likert scale (see below)

[14] The most common usage of ordinal scales is in obtaining attitude or preference measurements. The analyst is typically interested in whether something has more or less of a characteristic or attribute. Unlike an interval scale, the ordinal scale represents a ranking of an attribute because the intervals are not exactly equal for all the points on the scale. This limits what statistics can be used, making it impossible to determine how much more or less of an attribute something has. The analyst can only determine whether something has more or less of an attribute or characteristic, but not to what extent.

Questionnaire Item and Responses

Question:	A two-level maintenance concept can be used to maintain this system.				
Scale Descriptor:	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Assigned Value:	1	2	3	4	5
Number of Responses:	0	0	1	6	0
Results					
Mean Response: 3.8 ((1 x 3) + ((6 x 4))/7					

To rectify this situation, the analyst selects the mode as the metric (Table 1-21). Using the mode as the metric, the analyst rates the measure as "met criteria" because the mode value of 4 corresponds to "Somewhat Agree." This example highlights how using inappropriate metrics can sometimes produce different results.

Table 1-21: Appropriate Analysis of Ordinal Data Example

Measure Description

Task	Attribute	Measure	Metric	Criteria(Threshold)	Data
Maintain and Sustain System	Maintainability	Logistician rating of maintainability (MOS)	Mode	≥ Somewhat Agree	Logistician responses to questionnaire item with 5-point Likert scale (see below)

Questionnaire Item and Responses

Question:	A two-level maintenance concept can be used to maintain this system.				
Scale Descriptor:	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Assigned Value:	1	2	3	4	5
Number of Responses:	0	0	1	6	0
Results					
Mode Response: 4 (Somewhat Agree)					

1.11.3. Analyzing Data with "Not Applicable" or "Don't Know" Selections

When analyzing and presenting data for an item, the analyst should exclude "Not Applicable" and "Don't Know" selections from the other responses. Including "Not Applicable" or "Don't Know" selections with the other responses can produce misleading results and lead to incorrect interpretations of the data.

Figures 1-3 and 1-4 show how the response distributions change significantly when the "Don't Know" selections are inappropriately included as part of the data set.

In Figure 1-3 (without "Don't Know" selections), the majority of respondents (8 out of 10) who are qualified to answer the question think the reconnaissance map is fairly or very important for the ground planning mission. With the addition of the "Don' Know" selections in Figure 1-4, it appears there is no longer a majority of respondents who think the reconnaissance map is important for the ground planning mission. The visual image presented in the figure draws attention to the high number of respondents who did not have adequate knowledge or experience on which to base an answer. In this case, interpreting the results becomes more difficult and may lead to faulty conclusions. For instance, one may incorrectly conclude from the figure that less than half of the respondents think the airborne radiation survey reconnaissance map is important for the ground planning mission.

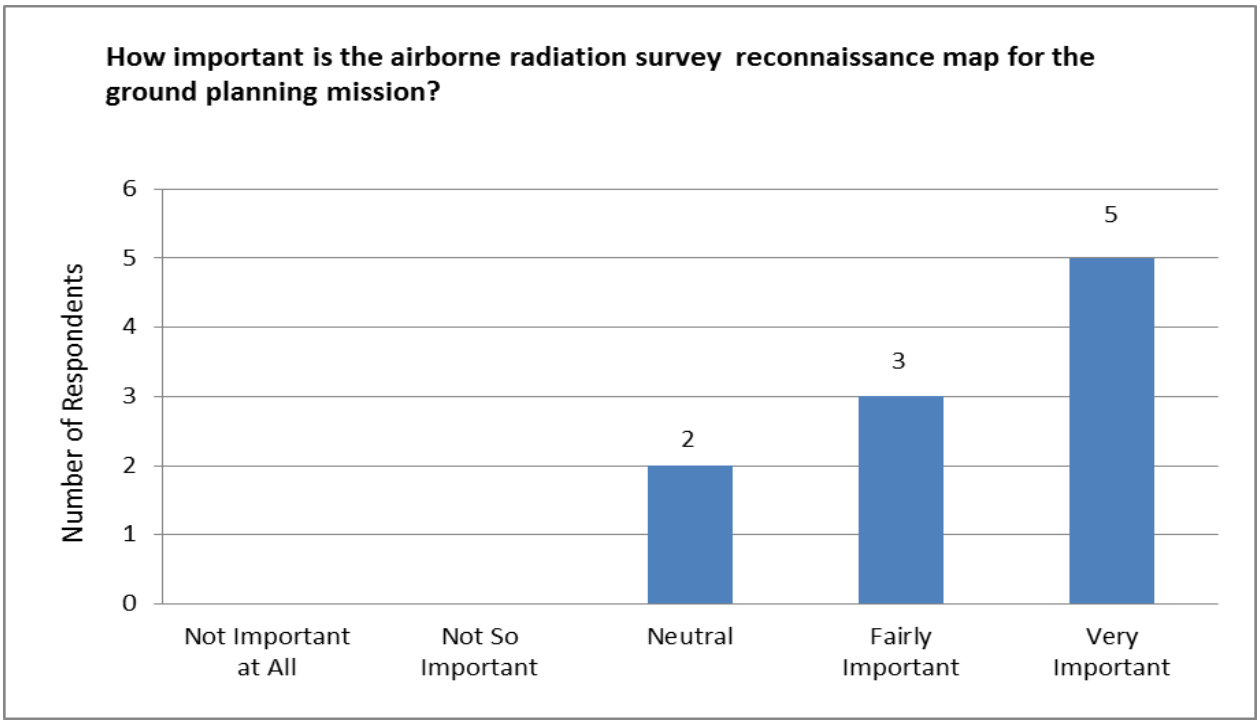


Figure 1-3: Bar Chart without "Don't Know" Selections

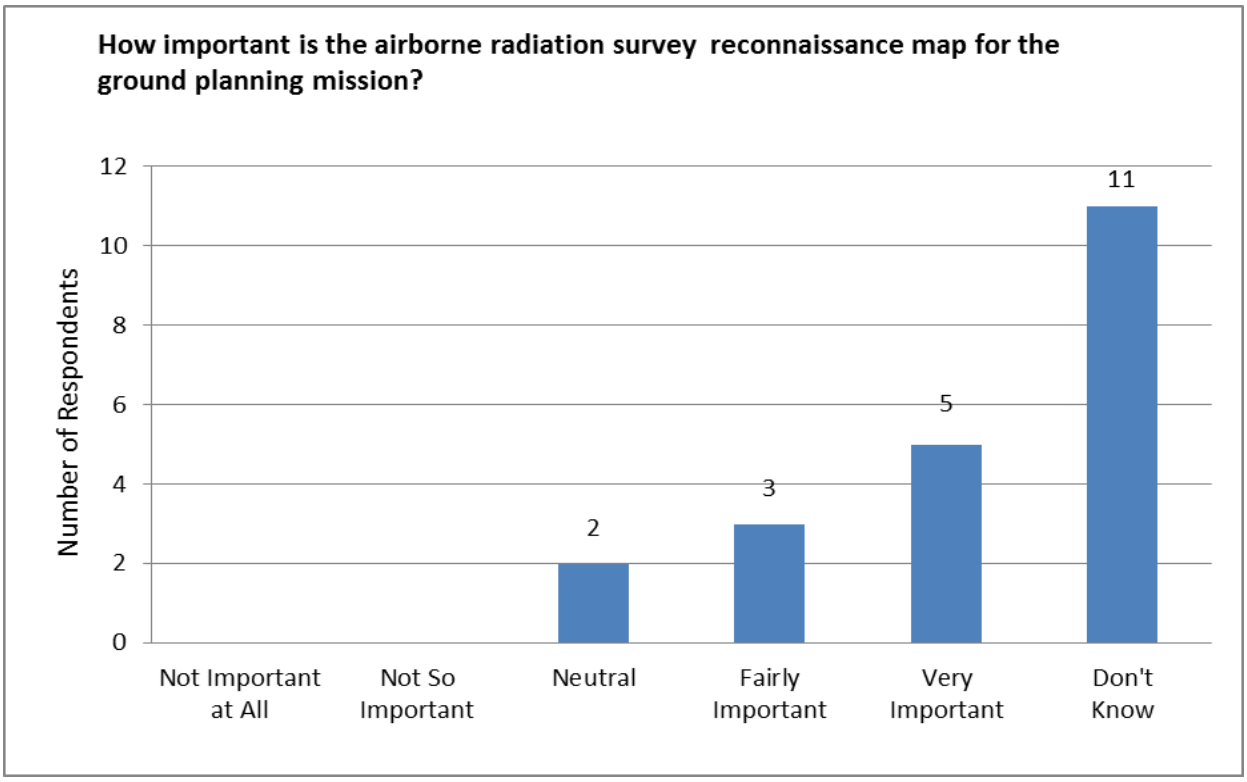


Figure 1-4: Bar Chart with "Don't Know" Selections

1.11.4. Sensitivity Analysis

Though there are numerous ways to analyze data and produce results, the sensitivity analysis is distinct in that it can yield new and meaningful insights that can profoundly influence the interpretation of the results. The main purpose of the sensitivity analysis is to highlight performance stability or robustness of the system, solution, or concept being assessed in the study. This is accomplished by varying performance parameters, operational conditions, scenarios, or assumptions to determine the resulting changes in performance. The sensitivity analysis not only enhances the credibility of the analysis, but also facilitates the identification of key performance tradeoffs. The results of this analysis often serve as the basis for study conclusions, recommendations, and decisions.

There are a number of ways data collected in surveys can be used in a sensitivity analysis. For example, if experts are asked to estimate a range of possible reliability values of a system, then the analyst could vary the reliability values within the specified range to determine whether the overall results significantly change. As another example, the analyst may want to identify which variables and associated point estimates collected through a survey significantly drive results. For these variables, the analyst could determine whether any reasonable changes in the point estimates produce very different results. The analyst could use such information to inform decision-makers of the potential risks associated with any decisions based on the results of the analysis.

2.0. EXPERT ELICITATION

After an introduction of expert elicitation and judgment, this chapter presents an approach to conducting expert elicitation in operational capability requirements studies. It provides insights into the selection of experts, development of questions, and design and conduct of the elicitation process.

2.1. Introduction

As a special form of survey research, expert elicitation relies on surveys to collect information from subject matter experts. It is a structured method of gathering expert judgment and answering questions concerning issues or problems of interest in a study. The Delphi method, developed by the RAND Corporation in the 1950s, was one of the first recognized expert elicitation methods.[15] Over the years, many other elicitation methods have been developed and used by various organizations in both the private and public sectors. There are numerous examples of its use by federal agencies to include the United States Army Corps of Engineers, Nuclear Regulatory Commission, National Aeronautics and Space Administration, Department of Transportation, Department of Energy, Department of Agriculture, and the Environmental Protection Agency.

Since expert judgment is affected by the approach used to gather it, a specially designed process is required that includes procedures for developing questions, conducting the elicitation, and handling biases that may arise. The process is designed to facilitate thinking and encourage experts to state their true opinions. Through the elicitation process, experts derive judgments from the available body of evidence ranging from direct empirical data to theory. Although the process is formal and structured, it can differ in terms of the degree of interaction between experts, level of detail in information elicited, number of meetings, type of communication mode, and degree of structure in the elicitation process.

Expert elicitation is different from probability sampling methods since respondents are not considered to be representative of a population (Chan et al, 2010). Instead, respondents are viewed as representing a large body of knowledge. Expert elicitation seeks to reflect the range of credible opinions regarding a specific question or problem, so the foremost concern is the quality and diversity of the participating experts.

2.2. What is an Expert?

Meyer and Booker (2001) define an expert as “a person who has background in the subject area and is recognized by his or her peers or those conducting the study as qualified to answer questions.” It is natural to think of experts as professionals such as scientists, physicians, and engineers, but any person with sufficient knowledge of the subject matter can be considered an expert for the purpose of the study. Although an individual’s knowledge is important, other factors such as personality, experience, and expertise in organizing and using his or her knowledge are critical to the success of the elicitation (O’Hagan et al, 2006, p. 27). Achieving a balanced and broad spectrum of viewpoints may require eliciting judgments from individuals with various backgrounds and degrees of expertise.

[15] The Delphi method entails a group of experts who anonymously reply to questionnaires and subsequently receive feedback in the form of a statistical representation of the "group response," after which the process repeats itself. The goal is to reduce the range of responses and arrive at something closer to expert consensus.

2.3. Expert Judgment

There is a variety of terms used to describe expert judgment such as expert opinion, subject matter expert assessment, subject matter expert analysis, subjective judgment, and expert knowledge.

Whatever it is called, expert judgment is the data given by an expert in response to a question and represents an expression of opinion based on knowledge and experience. Judgment is shaped by the expert's state of knowledge at the time of the response to the question, and because experts have different experiences and knowledge, their judgments can differ and change over time as new information is learned.

Expert judgment is commonly expressed in quantitative terms, although it is possible to obtain expert judgment in a variety of other non-numeric or qualitative forms. Some examples of information elicited from experts are shown in Table 2-1.

Table 2-1: Examples of Information Elicited from Experts

Quantitative	*Qualitative*
Probability of an occurrence of an event	Impact of a change
Probability of failure of a system	Risks and consequence of a decision
Estimates of ranges of uncertainty	Variables, assumptions, and data used in an analysis
Likelihood of a causal relationship	Elements needed for decision making
Allocation of funding	Failure causes, potential failures, potential solutions
Rating of the performance of a model	Methods to optimize performance

2.4. An Expert Elicitation Approach for Capability Requirements Studies

It is necessary to follow a formal and structured process to ensure the information elicited from experts is suitable for analysis. The following describes a seven-step approach to conducting expert elicitation in the CBA, AoA, or other type of capability requirements study. It provides guidelines for the selection and preparation of experts, development of questions, design and conduct of the elicitation process, and analysis and reporting of data.

2.4.1. Step 1: Identify the Need for Expert Elicitation

In conducting a capability requirements study, the analyst typically deals with many unknowns associated with new and complex concepts. Choosing the appropriate research methods to collect and analyze data is a primary concern. Study objectives, data accessibility, time and resource constraints, and available tools and techniques are some important factors that the analyst must consider when determining which research methods to use.

In most studies, there is a need to gather subjective information for various reasons. Expert elicitation is one of several techniques that can be used to gather such information. Some examples of using expert elicitation for this purpose include the following:

- Determining the priority or rank of capability gaps,
- Establishing study ground rules, constraints, and assumptions,

- Determining the scope, purpose, and objectives of a study,
- Identifying criteria (threshold and objective values) of measures,
- Determining the screening criteria for alternatives,
- Identifying and rating risks, probability of risk, and risk consequences,
- Identifying and prioritizing threats and targets,
- Selecting scenarios to be used in a study,
- Identifying potential materiel and non-materiel solutions,
- Determining the most viable alternatives.

In some cases, expert elicitation can be used when other research methods (e.g., modeling and simulation, parametric analysis) are not feasible or data is insufficient, unattainable, or too costly or impractical to collect. In these cases, the data and information elicited from experts is typically used in capability requirements studies to compute a metric for a measure of effectiveness, suitability, or performance.[16] For example, experts may be asked to provide estimates of a system's probability of kill performance against specific targets under certain operational conditions. The mean, median, or other appropriate metric for the measure is used to compute a value from the estimates. The analyst uses the value to determine whether the measure criteria are met or not. The data and information elicited from experts can address various attributes of interest in the study such as those shown in Table 2-2.

Table 2-2: Examples of Attributes

Accessibility	Accuracy	Adaptability	Adequacy
Availability	Capacity	Compatibility	Deployability
Flexibility	Interoperability	Latency	Maintainability
Mobility	Networkability	Persistence	Precision
Reliability	Scalability	Security	Simplicity
Sufficiency	Survivability	Sustainability	Time
Timeliness	Transportability	Utility	Vulnerability

2.4.2. Step 2: Develop the Questions

Expert elicitation relies on surveys to collect data of some aspect for analysis. Expert judgment is primarily elicited through face-to-face interviews. The choice of whether to use personal interviews (i.e., interview one expert at a time) or group interviews (i.e., interview experts in a group) will depend on various factors such as time constraints and the availability of experts. Whatever method is chosen, using good questions is an essential part of the survey process.

Crafting good questions requires careful forethought and a sound approach. The process entails drafting a set of initial questions and using a small group of experts to design the final questions. Subject matter experts who are not among the experts in the panel can assist in developing the questions and any assumptions, definitions, or other supporting information. Expert insights gleaned during the question development process will help ensure the questions are eliciting the information of interest in the study.

[16] For more information on metrics and measures, see *The Measures Handbook*, Office of Aerospace Studies, 6 August 2014.

Capability requirements studies typically require many different types of experts (e.g., aircraft operators, logisticians, intelligence experts), so it is critical to have the right ones participating at the right time.

Pre-testing the questions with several other experts can help refine the questions and identify problems such as unclear wording or misreading that must be addressed prior to using the questions in the elicitation. Feedback from experts will be helpful in determining how specific questions should be worded, order and number of questions, and question format. See Section 1.3 (Designing Good Questions) for additional information regarding question development.

2.4.3. Step 3: Select the Experts

Selection criteria define the set of individuals that have a chance of being selected to participate as expert panel members in the study. With the selection of experts being a critical step in the process, it is important to establish selection criteria through careful deliberation. Given that the expert panel selection is not random, there is a risk of researcher bias when the researcher makes selections based on inappropriate criteria. Selection error present in an expert panel depends on the degree of expertise of the person making the selection decision. It is advantageous to consider a range of possible criteria by drawing from the expertise of the study director, study team members, study advisory group, and other appropriate individuals and groups.

A “good” expert has technical knowledge, experience, and intuition as well as an ability to integrate information and draw conclusions. Criteria such as level of training, type of skill, and years of experience can be used to ensure the panel consists of experts with the proper knowledge and expertise. Ultimately, selection criteria will depend on the objectives of the study. Table 2-3 provides some examples of criteria that can be used to identify experts for participation in a study.

Like other studies, the number of experts used in a capability requirements study will be driven mostly by resources and time available to conduct the study as well as the number and availability of individuals who have the expertise in the area being studied. A major challenge for the analyst is not only identifying experts, but also getting them to participate. In some cases, there may be no experts. For example, if the area of interest involves break-through technologies or existing systems will be used differently from how they were designed, then there may well be no individuals with the appropriate experience and expertise to be deemed “experts”. In these cases, the analyst may still find the “most knowledgeable” individuals and elicit information from them, but the uncertainty of the information may be problematic for the study. Other research methods may be required in these cases. In other cases, experts may exist, but they may not be available to participate in a study because their time is in very high demand. Other individuals may be recommended, but they may not have the requisite experience and expertise in the area of interest. Similar to the situation described above, these cases may require the use of other research methods.

Table 2-3: Examples of Selection Criteria

Criteria	Description
Knowledge of Area of Interest	Understanding of the area of interest, reputation as a technical authority, awards received, membership in organizations or groups in the area of interest.
Background and Experience	Years of experience, level and diversity of experience, type and number of past positions held.
Education and Training	Specialized training, type of advanced academic degree(s), special certification(s) and qualifications.
Published Work	Number and quality of publications in the area of interest.
Personal Skills	Interpersonal skills, communication skills, flexibility, impartiality, ability to generalize and simplify.
Economic or Personal Stake	Lack of economic or personal stake in the potential findings.
Availability and Willingness	Availability and willingness to commit the necessary time and effort to participate in the study, willingness to prepare for discussions and provide opinions.

Although there are no absolute rules regarding the number of experts, large panels increase the likelihood that all possible expert views are represented. While all are knowledgeable of the area of interest, experts have different experiences and perspectives that will shape their responses. Large panels can often produce insights that may not be possible with small panels.

Despite the lack of definitive approaches to determining the appropriate number of experts, a panel of practitioners in expert elicitation recommends at least six experts should be included and that the benefit of including additional experts beyond 12 begins to diminish (Cooke and Probst, 2006, p. 16). Using panels with less than six members will likely reduce the chances of collecting a diversity of information.

2.4.4. Step 4: Prepare the Experts

Once the experts have been identified and selected, the next step entails preparing them for the elicitation by providing relevant information about the study. Experts must have a thorough understanding of the issues before they are ready to answer questions. Issue familiarization is the process used to help the experts understand the issues of interest in the study, purpose of their participation, expectations, study objectives, elicitation process, list of questions, terminology, and key assumptions and definitions. Depending on the objectives of the elicitation, information about the technical aspects of the baseline capabilities, potential solutions, study methodology, and measures (e.g., MOEs, MOSs, and MOPs) may be required as well. Whether done in a group or individually, it is important to present the same information to ensure a common understanding of the issues. Presentations, briefing books, and other documents should be assembled to provide the relevant information.

2.4.5. Step 5: Conduct the Elicitation

The approaches used to elicit judgments vary widely and will rely to a large degree on the objectives of the study. The amount of time required for the elicitation may range from a few hours to as much as a week depending on the size and complexity of the study. The analyst should consider a number of factors in designing the elicitation:

- Time and resources available for the study,

- Type of information to be elicited,
- Number of experts,
- Amount of time experts will need to provide judgments,
- Degree of interaction among the experts,
- Number and type of questions,
- Format for the answers,
- Mode(s) of communication,
- Type of interview.

Expert judgment is elicited through personal interviews, group interviews, or a combination of both. Personal interviews are usually done in private and in person and allow the interviewer to gather in- depth data from the experts without distraction or influence by other experts. Group interviews are conducted in person through a structured approach that defines how experts express and discuss their opinions.

Although personal interviews can be used, convening an in-person group meeting to conduct the elicitation has several advantages in the CBA, AoA, and other capability requirements study. Most importantly, it provides an opportunity to introduce the issue, review the relevant information, and describe the elicitation purpose and process. It can serve as a forum to answer questions, share information, discuss expectations, describe how the results will be used, and gain feedback on any issues that require further clarification or additional information. The major drawback to group elicitation is the undesirable effects of dominant or vocal participants, something that is avoided by eliciting experts individually through personal interviews (Cooke and Probst, 2006, p. 16).

In group elicitations, there are greater demands of time and effort on the interviewer to structure and facilitate the discussions and interactions amongst the experts. The interviewer is responsible for ensuring the integrity of the elicitation process and its implementation by initiating and maintaining effective discussions. Ayyub (2001, p. 18) recommends using a facilitator or moderator to help create an environment that ensures equity in presenting views and a successful elicitation of opinions and information from each expert.

In capability requirements studies, gaining insights into the underlying reasoning or rationale of an expert's response may be as important as the response itself. There are several techniques described by Meyer and Booker (2001) that can be used to interview experts and learn the rationale for a response:

- The verbal report involves instructing the expert to think aloud when answering a question and resembles someone talking to oneself. The technique can be time consuming since it is used on one expert at a time. It is important to note that not all experts are capable of verbalizing all their thoughts for various reasons (e.g., too difficult to articulate, thoughts are automatic or unconscious).
- The verbal probe entails phrasing questions in a way to minimize influencing the expert's thinking. The technique is a quick means of obtaining information and is suitable for both personal and group interviews. Some examples include repeating the question, asking "Anything else?", "Tell me more", and "How do you mean that?"
- The ethnographic technique involves transposing the expert's words into questions. Because the questions are based on the expert's own words, it is a non-biasing form of questioning. The technique can be time consuming and is not suitable for group interviews.

In structuring the elicitation, it is important to understand and anticipate bias that may occur. Bias is a skewing that arises from our personal perceptions and understanding. There are various forms of bias and methods for dealing with them. Table 2-4 provides a brief description of seven common forms of bias and when they are likely to occur.

Several steps can be taken in designing the elicitation process to help mitigate anticipated bias. For example, to reduce social pressure from the data gatherer, the interviewer can use the verbal report, verbal probe, and/or ethnographic phrasing of questions instead of direct questions that may lead the experts. If complicated response forms such as probability and uncertainty estimates are being elicited, prepare the experts for the elicitation by conducting a training session that describes the fundamental principles of the response form. The training will help eliminate the potential of confusion and underestimation and give the experts an opportunity to rehearse providing responses to sample questions in the appropriate form. Finally, as part of the preparation for the elicitation, it is important to make the experts aware of the forms of bias and why they happen. Although bias cannot be completely eliminated, experts will not be able to control their own tendencies toward bias without first having a good understanding of it.

While much can be done to design the elicitation to help mitigate bias, the interviewer must still be alert to the occurrences of bias during the elicitation process and make the appropriate adjustments to counter it. For example, if there are inconsistencies in responses, the interviewer should ask the experts to reconsider their responses. If fatigue is a factor, the interviewer can shorten the elicitation sessions or schedule breaks to help preclude potential inconsistencies in responses. In group situations, the interviewer should suspect group think is occurring when no one in the group voices a difference of opinion or the group consistently defers to one expert or a subgroup of experts in the group.

Table 2-4: Common Forms of Bias

Bias	Description
Social Pressure (Data Gatherer)	Individuals consciously or unconsciously alter the descriptions of their thoughts to gain acceptance and to be seen in the most positive light possible. Data gatherers can intentionally or unintentionally influence the individual through body language, facial expression, intonation, and word choice. More pronounced in cases when the interviewer uses leading questions.
Social Pressure (Group Think)	Social pressure from others in a group induces individuals to alter their responses or silently acquiesce to what they believe will be acceptable to the group. More pronounced when individuals in a group desire to remain as members, are satisfied with the group, and view the group as cohesive.
Wishful Thinking	Individuals' hopes influence their judgment—what individuals think should happen will influence what they think will happen. More pronounced when individuals do not have to explain their reasoning and when individuals are personally involved or would gain from their answers.

Bias	Description
Inconsistency	Individuals are inconsistent in solving problems—their current thoughts or answers may contradict those expressed earlier. More pronounced when: 1. Elicitation sessions are long and individuals forget instructions, definitions, and assumptions, 2. Complicated response forms such as probability distributions and percentiles are causing confusion, 3. Experts are asked to consider too many things and become confused and inconsistent.
Underestimation of Uncertainty	Individuals underestimate the uncertainty in the answers they provide. More pronounced when response forms are probabilities and other quantitative estimates.
Anchoring	Individuals receive additional information but do not adjust from their first impression in answering the question. More pronounced when experts have described their positions orally or in writing and fear losing face if they change their response.
Availability	Individuals do not mention more than one or two considerations in giving their responses which can mean the experts are drawing from data that is easier to recall. More pronounced when the expert does not receive any information from others that could help trigger less accessible data when formulating a response.
Source: Adapted from Meyer and Booker, 2001, p. 133	

There are many different approaches to interview experts that would be appropriate in these studies. In group situations, one approach commonly used involves interviewing each expert separately, reviewing the answers in a group, and then providing an opportunity for the experts to revise their responses.

Depending on the objectives of the study, the analyst may only be interested in collecting responses to questions, whereas in other cases, the rationale for the response may be required as well. Table 2-5 shows several examples of elicitation methods for group interview situations.

Table 2-5: Examples of Elicitation Methods for Group Interview Situations

Example A
Each expert in a panel is asked individually to provide a response to a question as well as rationale for his or her response that includes identification of issues that significantly influenced the response. After providing responses, the panel of experts is given an opportunity to review the results. During the review, each expert discusses the rationale for his or her response while the other panel members are encouraged to ask questions and contribute information. Following the review, the experts are given an opportunity to revise their responses and provide rationale in light of what was learned during the discussion. With the submission of the revised responses, the question is closed and the elicitation process resumes with the next question.
Example B
Each expert is asked individually to provide an initial response to a question. To avoid social pressure, the individual responses are then displayed anonymously to the panel of experts through an on-screen graphical presentation. The experts are given an opportunity to discuss the results of the presentation. Following the discussion, each expert is asked individually to provide a final response. With the submission of the final response, the question is closed and the elicitation resumes with the next question.
Example C
Questions with associated background information are provided to the panel of experts. To encourage knowledge sharing, the experts are given an opportunity to discuss the questions and information as a group. The interviewer monitors the discussion and responds to any questions from the panel members. If necessary, the interviewer provides additional information to help the panel in understanding the issues. The information may be requested by the panel, or the interviewer, through observation, deems the information is needed to facilitate the discussion. When the panel discussion is complete, each expert is asked individually to provide a response to each question. With the submission of the responses, the questions are closed and the elicitation resumes with the next set of questions.

In personal interview situations, experts are interviewed separately in face-to-face meetings or by telephone. If the response requires clarification or there is a desire to collect the rationale for the response, the analyst can use the verbal report, verbal probe, or ethnocentric technique described earlier to gather the information. For example, an analyst can instruct the experts to explain in detail their thinking process as they respond to the questions (verbal report). The verbal probe and ethnographic technique can be used to clarify responses and/or gain more insights into the rationale for the responses.

The questions used in the elicitation will depend on the objectives of the capability requirements study. Questions can be designed to elicit opinions in a variety of forms such as quantities, uncertainties, relationships, parameters, or events. Table 2-6 shows several examples of information that can be elicited.

Table 2-6: Examples of Information Elicited from Experts

Example A
In determining the probability of a system failure, experts are asked to provide a best estimate as well as a degree of uncertainty. The best estimate is expressed as a percentage, although the decimal or ratio can be used as well. This estimate is viewed as the median value where there is a 50% chance that the "true" value will be higher, and a 50% chance the "true" value will be lower. Next, the experts are asked to estimate an upper bound where there is a strong likelihood (95% chance) that the "true" value will be lower than the estimate, and only 5% chance that the "true" value will be higher. In the analysis, these estimates are used as the 50^{th} and 95^{th} percentile values.
Example B
After reviewing technical information of a system, the experts are asked to rate how easily the system can be configured for transport. Each expert is asked to answer a series of questions with five-point Likert scales ranging from "strongly disagree" to "strongly agree" and provide written rationale for his or her response. In the analysis, the mode statistic is determined for each question and the rationale used by the experts is highlighted in the discussion of the results.
Example C
Experts are given an opportunity to review five models used for predicting performance of a system. Each expert is asked to rate the plausibility of each model using a seven-point scale ranging from "Least Plausible" to "Most Plausible" and provide written rationale for his or her response. In the analysis, the responses are shown graphically along with the mode statistic for each model. The results provide a discussion of the mode statistic and rationale used by the experts in rating the models.

2.4.6. Step 6: Aggregate the Data

In capability requirements studies, there is typically a requirement to report a single value by combining responses. Whether judgments are elicited from experts separately or in a group in some instances, the analyst can mathematically aggregate the responses using simple algorithms such as the mean and median. For example, if experts were asked to provide an estimate of a system's reliability (a probability value), the analyst can use the mean, median, or other simple algorithms to aggregate the estimates.

More complex weighted means can be used to give more weight to experts who are viewed as having more expertise, although the prevailing recommendation among practitioners in expert elicitation is to use equal weights since it is a simple and robust method for aggregating expert judgments (O'Hagan, 2006, p. 222; Meyer and Booker, 2001, p. 329). Measurement scales such as the Likert scale produce ordinal data, so it is important to use appropriate statistics such as the mode or median.

If the judgments are elicited from experts in a group, another option is to use a behavioral aggregation that requires a convergence or consensus of opinion among the experts through discussion and interaction. A major risk of this approach is the undue influence of dominant participants.

2.4.7. Step 7: Report the Results

Since there is both potential value and danger of using expert judgment, some guidelines are necessary when reporting results derived from expert judgment. Traditional scientific research does not explicitly accommodate the use of opinions as scientific data. It is important to ensure the distinction between

empirical data and expert judgment data is maintained by clearly identifying which analyses are based on empirical data and which are based on expert judgment data. Cooke (1991) recommends that sufficient information should be provided about the data and calculations so that the results can be reproduced by others.

Another important consideration is the generalizability of results. Unlike probability sampling, expert elicitation is unlikely to produce results that are representative of a population since all individuals in the population do not have equal chances of being selected. This means the analyst should not make statistical inferences about a population from the expert judgment data. Expert elicitation does not entail randomly selecting individuals with the intent of making inferences about a population, but rather, individuals are selected based on their knowledge and experience with the intent of drawing conclusions about the existing knowledge base. Lastly, the analyst should provide the names and background information of the experts used in the study in the final report. This will help readers ascertain the credibility of the experts.

2.5. Summary

Expert elicitation can be a useful technique for gathering various types of data for analysis in the CBA, AoA, and other capability requirements study. Expert elicitation is a formal and structured process that entails the selection of experts, conduct of the elicitation, and analysis of data. The approach described in this chapter will help ensure the information elicited from experts is properly collected and suitable for analysis. It provides guidelines for the selection and preparation of experts, development of questions, design and conduct of the elicitation process, and analysis and reporting of data.

ANNEX A: Acronyms

ANOVA	Analysis of Variance
AoA	Analysis of Alternatives
CBA	Capabilities-Based Assessment
IQR	Interquartile Range
JCIDS	Joint Capabilities Integration and Development System
JS	Joint Staff
MOE	Measure of Effectiveness
MOP	Measure of Performance
MOS	Measure of Suitability
OAS	Office of Aerospace Studies

ANNEX B: Information Sources

A5/7 Capability Development Guidebook, Volume 2D, Annex A, Analysis of Alternatives

A5/7 Capability Development Guidebook, Volume 2C, Capability Based Assessment

Defense Acquisition University (DAU) Glossary

Manual for the Operation of the Joint Capabilities Integration and Development System (JCIDS Manual)

ANNEX C: References

Ayyub, Bilal M. (2001). *A Practical Guide on Conducting Expert-Opinion Elicitation of Probabilities and Consequences for the Corps Facilities*. Alexandria, VA: Institute for Water Resources Report 01-R-01, United States Army Corps of Engineers.

Chan, Gabriel, Laura D. Anadon, Melissa Chan, Audrey Lee. (2010). Expert Elicitation of Cost, Performance, and R&D Budgets for Coal Power with Carbon Capture and Sequestration. *Energy Procedia*.

Churchill, Jr., Gilbert A. (1979). *Marketing Research: Methodological Foundations, Second Edition*. Hinsdale, Illinois: The Dryden Press.

Cooke, Roger M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* New York, NY: Oxford University Press.

Cooke, Roger M. and Katherine N. Probst. (2006). *Highlights of the Expert Judgment Policy Symposium and Technical Workshop*. Washington DC.

DeVellis, Robert F. (1991). *Scale Development: Theory and Applications*. Newbury Park, CA: Sage Publications.

Emory, C. William. (1985). *Business Research Methods, 3rd Edition*. Homewood, IL: Richard D. Irwin, Inc.

Fink, Arlene. (2003). *The Survey Handbook, Second Edition*. Thousand Oaks, CA: Sage Publications, Inc.

Fowler, Jr., John J. (1993). *Survey Research Methods, Second Edition*. Newbury Park, CA: Sage Publications, Inc.

Kerlinger, Fred N. (1986). *Foundations of Behavioral Research, Third Edition*. Fort Worth, Texas: Harcourt Brace Jovanovich College Publishers.

Leedy, Paul D. (1997). *Practical Research: Planning and Design, Sixth Edition*. Upper Saddle River, New Jersey: Prentice-Hall, Inc.

Meyer, M. A. and J.M. Booker. (2001). *Eliciting and Analyzing Expert Judgment: A Practical Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics, American Statistical Association.

O'Brien, Thomas G. and Samuel G. Charlton (editors). (1996). *Handbook of Human Factors Testing and Evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. (2001). *Uncertain Judgements: Eliciting Experts Probabilities*. West Sussex, England: John Wiley and Sons, Ltd.

Tull, Donald S. and Del I. Hawkins. (1980). *Marketing Research: Measurement and Method*. New York, New York: Macmillan Publishing Co., Inc.

Zikmund, William G. (1991). *Exploring Marketing Research, 4th Edition*. Hinsdale, IL: The Dryden Press

ANNEX D: Glossary

Attribute – a quality or feature of something. Attributes of tasks (e.g., survivability, persistence, availability, accuracy, etc.) form the basis for identifying and drafting measures. (AFOTECMAN 99-101)

Capability – the ability to complete a task or execute a course of action under specified conditions and level of performance. (JCIDS Manual 2015)

Capability Gap (or Gap) – the inability to meet or exceed a capability requirement, resulting in an associated risk until closed or mitigated. The gap may be the result of no fielded capability, lack of proficiency or sufficiency in a fielded capability solution, or the need to replace a fielded capability solution to prevent a future gap. (JCIDS Manual 2015)

Capability Requirement – a capability which is required to meet an organization's roles, functions, and missions in current or future operations. To the greatest extent possible, capability requirements are described in relation to tasks, standards, and conditions in accordance with the Universal Joint Task List or equivalent DOD Component Task List. If a capability requirement is not satisfied by a capability solution, then there is also an associated capability gap which carries a certain amount of risk until eliminated. A requirement is "draft" or "proposed" until validated by the appropriate authority. (JCIDS Manual 2015)

Confidence Interval – a specified range of numbers within which a population parameter should lie; an estimate of the population parameter based on the knowledge that it will equate the sample statistic plus or minus a sampling error. (Zikmund 1991)

Confidence Level – a percentage or decimal value that tells how confident a researcher can be about being correct. It states the long-run percentage of confidence intervals, including the true population mean. (Zikmund 1991)

Criteria (also referred to as Standards) – define the acceptable levels or standards of performance for a metric and are often expressed as a minimum acceptable level of performance (threshold) and desired acceptable level of performance (objective). (AFOTEC)

Data – individual measurements that are used to compute the metric for a measure. (AFOTEC)

Measure – a device designed to convey information about an entity being addressed. It is the dimensions, capacity, or amount of an attribute an entity possesses. (AFOTEC)

Measure of Effectiveness – a measure designed to correspond to accomplishment of mission objectives and achievement of desired results. (DAU Glossary)

Measure of Performance – a measure of a system's performance expressed as speed, payload, range, time on station, frequency, or other distinctly quantifiable performance features. (DAU Glossary)

Measure of Suitability – a measure of an item's ability to be supported in its intended operational environment. (DAU Glossary)

Metric – a unit of measure that coincides with a specific method, procedure, or analysis (e.g., function or algorithm). Examples include: mean, median, mode, percentage, and percentile. (AFOTEC)

Objective – an operationally significant increment above the threshold. An objective value may be the same as the threshold value when an operationally significant increment above the threshold is not identifiable.

Sample – a portion or subset of a larger group called a population. (Fink 2003)

Survey – a system for collecting information from or about people in order to describe, compare, or explain their knowledge, attitudes, and behavior. (Fink 2003)

Threshold – a minimum acceptable operational value of a system capability or characteristic below which the utility of the system becomes questionable

.

ANNEX E: Examples of Five, Six, and Seven Point Scale Descriptors

Five Point Scales				
Totally Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Totally Adequate
Very Inadequate	Slightly Inadequate	Borderline	Slightly Adequate	Very Adequate
Completely Unacceptable	Somewhat Unacceptable	Borderline	Somewhat Acceptable	Completely Acceptable
Largely Unacceptable	Barely Unacceptable	Borderline	Barely Acceptable	Largely Acceptable
Completely Ineffective	Somewhat Ineffective	Borderline	Somewhat Effective	Completely Effective
Very Ineffective	Ineffective	Borderline	Effective	Very Effective
Extremely Difficult	Somewhat Difficult	Borderline	Somewhat Easy	Extremely Easy
Completely Disagree	Substantially Disagree	Borderline	Substantially Agree	Completely Agree
Extremely Unimportant	Moderately Unimportant	Borderline	Moderately Important	Extremely Important
Completely Useless	Somewhat Useless	Borderline	Somewhat Useful	Completely Useful
Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Undoubtedly Worse	Moderately Worse	The Same	Moderately Better	Undoubtedly Better
Undoubtedly Worst	Noticeably Worse	Borderline	Moderately Better	Undoubtedly Best
Strongly Dislike	Dislike	Neutral	Like	Strongly Like
Never	Rarely	Now and Then	Often	Always
Never	Seldom	Now and Then	Frequently	Always
Poor	Fair	Good	Very Good	Excellent
Not at all Important	Not so Important	Neutral	Fairly Important	Very Important
Very Unimportant	Not Important	Borderline	Important	Very Important
Completely Dissatisfied	Somewhat Dissatisfied	Neither Satisfied Nor Dissatisfied	Somewhat Satisfied	Completely Satisfied
Hardly ever	Sometimes	Often	Very Often	All the Time
Never	Rarely	Sometimes	Often	Very Often
Not at all Unique	Slightly Unique	Somewhat Unique	Very Unique	Extremely Unique

Six Point Scales					
Totally Inadequate	Very Inadequate	Somewhat Inadequate	Somewhat Adequate	Very Adequate	Totally Adequate
Completely Unacceptable	Largely Unacceptable	Somewhat Unacceptable	Somewhat Acceptable	Largely Acceptable	Completely Acceptable
Completely Ineffective	Largely Ineffective	Somewhat Ineffective	Somewhat Effective	Largely Effective	Completely Effective
Extremely Difficult	Moderately Difficult	Somewhat Difficult	Somewhat Easy	Moderately Easy	Extremely Easy
Extremely Unimportant	Moderately Unimportant	Barely Unimportant	Barely Important	Moderately Important	Extremely Important
Completely Useless	Largely Useless	Somewhat Useless	Somewhat Useful	Largely Useful	Completely Useful
Completely Disagree	Substantially Disagree	Slightly Disagree	Slightly Agree	Substantially Agree	Completely Agree
Undoubtedly Worse	Moderately Worse	Slightly Worse	Slightly Better	Moderately Better	Undoubtedly Better
Never	Very Rarely	Somewhat Rarely	Somewhat Often	Very Often	Always

Seven Point Scales						
Totally Inadequate	Very Inadequate	Somewhat Inadequate	Borderline	Somewhat Adequate	Very Adequate	Totally Adequate
Totally Inadequate	Very Inadequate	Barely Inadequate	Borderline	Barely Adequate	Very Adequate	Totally Adequate
Completely Unacceptable	Largely Unacceptable	Somewhat Unacceptable	Borderline	Somewhat Acceptable	Largely Acceptable	Completely Acceptable
Completely Unacceptable	Moderately Unacceptable	Barely Unacceptable	Borderline	Barely Acceptable	Reasonably Acceptable	Completely Acceptable
Completely Ineffective	Largely Ineffective	Somewhat Ineffective	Borderline	Somewhat Effective	Largely Effective	Completely Effective
Completely Ineffective	Very Ineffective	Ineffective	Borderline	Effective	Very Effective	Completely Effective
Undoubtedly Worst	Conspicuously Worse	Moderately Worse	Alike	Moderately Better	Conspicuously Better	Undoubtedly Best
Extremely Difficult	Moderately Difficult	Somewhat Difficult	Borderline	Somewhat Easy	Moderately Easy	Extremely Easy
Completely Disagree	Substantially Disagree	Slightly Disagree	Borderline	Slightly Agree	Substantially Agree	Completely Agree
Extremely Unimportant	Moderately Unimportant	Barely Unimportant	Borderline	Barely Important	Moderately Important	Extremely Important
Completely Useless	Largely Useless	Somewhat Useless	Borderline	Somewhat Useful	Largely Useful	Completely Useful
Undoubtedly Worse	Moderately Worse	Slightly Worse	The Same	Slightly Better	Moderately Better	Undoubtedly Better
Never	Very Rarely	Somewhat Rarely	Borderline	Somewhat Often	Very Often	Always

ANNEX F: Determining Sample Size Involving Means

This appendix provides two versions of the equation that would be used in determining the sample size involving means. Version 1 is the equation without a finite population adjustment factor, whereas Version 2 is the equation with a finite population adjustment factor. When the size of a sample exceeds 5 percent of the population, it is acceptable to recognize that the finite limits of the population put a constraint on the size of the sample needed (Emory 1985, p. 296).

Unlike the equation used to determine sample size involving proportions described in Section 1.84 (Determining the Sample Size for Probability Sampling), these equations require a variance estimate (either the sample standard deviation or estimate of the population standard deviation). Previous studies or a pilot study may be used to estimate the standard deviation. If information is not available to estimate the standard deviation, a rule of thumb is to expect the standard deviation to be one-sixth of the range (Zikmund 1991, p. 507). For example, if the time required to perform routine maintenance on a new electronic warfare system is expected to range from 1 to 7 hours, then the rule of thumb estimate of the standard deviation would be 1 hour.

The standard error of the mean ($\sigma_{\bar{x}}$) is calculated in the same way as the standard error of the proportion (σ_p) that is described in Section 1.84. The analyst specifies a confidence interval (e.g., ± 0.1) and confidence level (e.g., .95). The z-value associated with a .95 confidence level is 1.96 standard errors. One standard error of the mean ($\sigma_{\bar{x}}$) is determined by dividing the confidence interval by the standard errors (0.1/1.96 = 0.051).

With the standard error of the mean ($\sigma_{\bar{x}}$) and standard deviation (*s*) determined, the analyst simply solves the equation for *n* to determine the sample size.

Version 1

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Solving for *n*, the equation becomes:

$$n = \frac{s^2}{\sigma_{\bar{x}}^2}$$

Where:

$\sigma_{\bar{x}}$ = standard error of the mean

s = sample standard deviation or estimate of the population standard deviation

n = size of the sample

Version 2

$$\sigma_{\bar{x}} = \frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}} \times \sqrt{\frac{\boldsymbol{N-n}}{\boldsymbol{N-1}}}$$

Solving for *n*, the equation becomes:

$$n = \frac{N\left(\frac{s}{\sigma_{\bar{x}} \times \sqrt{N-1}}\right)^2}{1 + \left(\frac{s}{\sigma_{\bar{x}} \times \sqrt{N-1}}\right)^2}$$

Where:

$\sigma_{\overline{x}}$ = standard error of the mean

s = sample standard deviation or estimate of the population standard deviation

n = size of the sample

N = size of the population