The Measures Handbook

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

6 August 2014

Office of Aerospace Studies

Air Force Materiel Command (AFMC) OAS/A5
1655 1st Street SE
Kirtland AFB, NM 87117-5522
(505) 846-8322, DSN 246-8322

DISTRIBUTION A. Approved for public release: distribution unlimited.
377ABW-2014-0666

Preface

The purpose of this handbook is to guide analysts in developing and analyzing measures for the Capabilities-Based Assessment (CBA), pre-Materiel Development Decision (MDD) analysis, and Analysis of Alternatives (AoA). The handbook is designed to supplement our CBA, pre-MDD analysis, and AoA handbooks by providing more detailed measure development and analysis guidance.

Given the importance of having good measures in these analyses, the handbook 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 his or her needs and the requirements of the study. The handbook also provides insights into various data collection and analysis methods that can be used to address measures in a study. Finally, guidance on interpreting and communicating results is provided to help the analyst construct a credible and defensible assessment of the results.

This handbook is grounded in over twenty years of experience in providing analytic guidance to Air Force and Department of Defense study teams and has been shaped by what OAS has learned and what we understand to be the expectations of senior decision-makers. As these lessons and expectations evolve, so will this handbook. If you have questions regarding specific information in the handbook or you have suggestions for improvements, please contact OAS at (OAS.DIR@us.af.mil) or 505-846-8322 (DSN 246).

OAS promotes an effective dialogue with the analyst community involved in planning and conducting CBAs, pre-MDD analyses, and AoAs. We encourage you to contact us and ask questions, or to provide always-appreciated feedback.

Jeff Erikson

Director, Office of Aerospace Studies

Table of Contents

List of Figures

List of Tables

1 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 at the initial Requirements Strategy Review (RSR) 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[1]) should be initiated. Furthermore, measures in all of these studies can be used to develop capability requirements in a Capability Development Document (CDD) and Capability Production Document (CPD). 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 (AFOTECMAN 99-101). 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 (DAU Glossary). 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:

[1] DOTmLPF-P acronym means: Doctrine, Organization, Training, materiel, Leadership and Education, Personnel, Facilities, and Policy.

- 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 (JCIDS Manual). The MOE is formally defined as:

> *A measure designed to correspond to accomplishment of mission objectives and achievement of desired results*. (DAU Glossary[2])

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 formally defined as:

> *A measure of an item's ability to be supported in its intended operational environment.* (DAU Glossary)

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

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

Note on Measure Labels

It is important to note that the labels used to describe types of measures are <u>not</u> critically important. In the CBA, for example, the generic term "measure" is frequently used without designating whether the measure is an MOE, MOS, or MOP. It is more important that the analyst focus on crafting good measures, instead of spending too much time trying to decide on the label to use.

[2] The definitions of MOEs, MOSs, and MOPs have subtle differences depending on the reference document used. To maintain consistency with 5000 series documents and the Joint Capabilities Integration and Development System (JCIDS), the DAU glossary version of the definitions is provided.

2 Understanding Measures and Data

This chapter presents several topics regarding measures and data. The chapter begins with a discussion of the four levels of measurement and the statistics that are appropriate for each level. This is followed by an explanation of the measure description elements and measure criteria. Next, the relationship between measures and high interest parameters and attributes is explained. Finally, the factors the analyst should consider when determining how to use data is discussed.

2.1 Levels of Measurement

As shown in Table 2-1, 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). Data associated with nominal and ordinal levels of measurement are commonly referred to as qualitative data, while 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 have an understanding of the levels of measurement to ensure the appropriate statistics are used.

Note on Analyst

The term "analyst" is used throughout this handbook and refers to one or more individuals or members of a team who are assigned to conduct all or some aspect 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 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 test, McNemar test, 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 – 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 (ANOVA), rank-order correlation
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, 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

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 (Universal Joint Task List Manual). 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	90^{th} Percentile	Threshold: $\leq$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.* (AFOTECMAN 99-101)
>
> *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).* (AFOTECMAN 99-101)
>
> *Data: individual measurements that are used to compute the metric for a measure.* (AFOTECMAN 99-101)
>
> *Conditions: describe the operational environment in which the task will be performed.* (Universal Joint Task List Manual)

2.3 Measure Criteria

Measure criteria[3] 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.* (AFI 10-601)
>
> *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.* (AFI 10-601)

[3] Criteria are also referred to as standards or standards of performance.

It is important to note that 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 CPD. Threshold and objective values used with measures developed in the CBA, pre-MDD analysis, and AoA can serve as the basis for developing threshold and objective values for capability requirements in the CDD and CPD.

In the ICD, capability requirements do not specify a threshold or objective value, but instead specify an initial objective value (formerly known as a minimum value) (draft JCIDS Manual 2014). 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 CPD). 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 CPD. 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 2012)
>
> *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 2012)

For all capability solutions being developed, there are five mandatory KPPs (i.e., Force Protection, System Survivability, Sustainment, Net Ready, and Energy Sustainment) that must be addressed by the study sponsor whether relevant to the capability or not (draft JCIDS Manual 2014). 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 CPD 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 CPD). An ICD specifies one or more capability requirements and associated capability gaps which represent unacceptable operational risk if left unmitigated. In a CDD and CPD, 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.

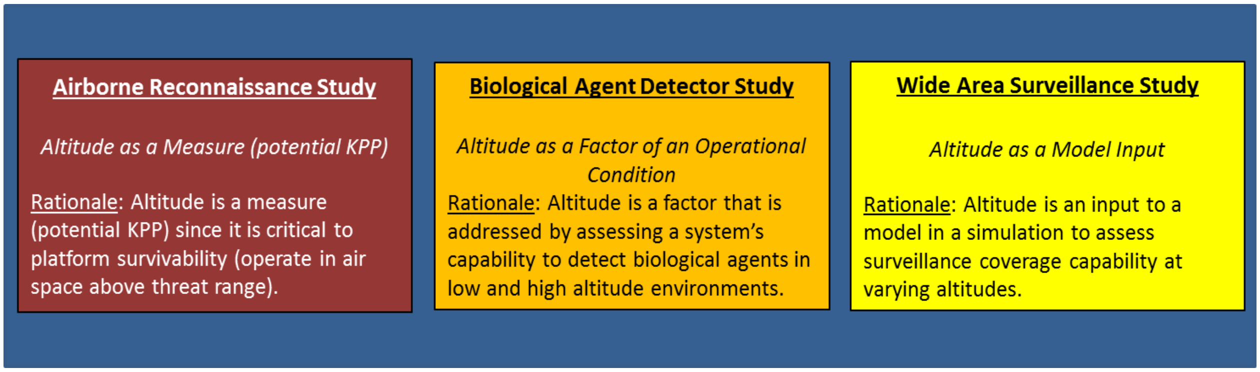


Figure 2-1: Examples of Using Data for Different Purposes

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 Measure Development Process

This chapter describes a four-step measure development process the analyst can use to develop measures for a study.

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.

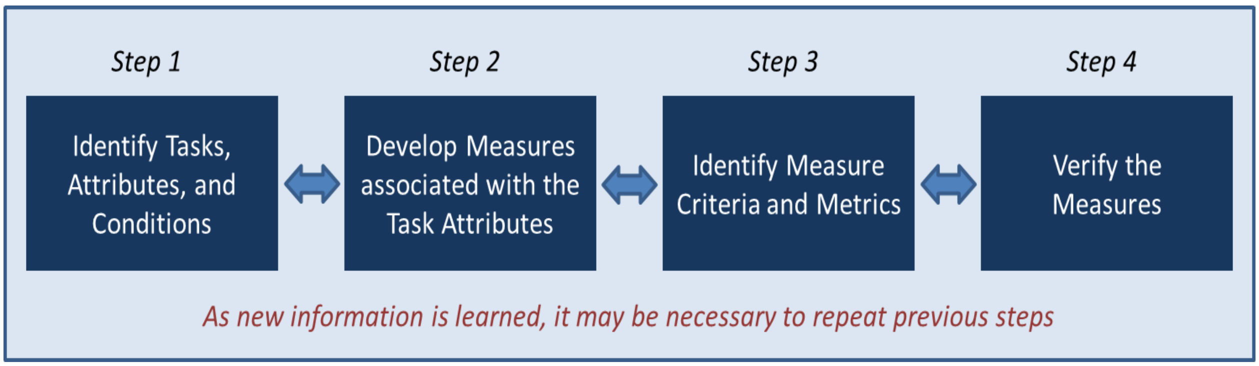


Figure 3-1: Measure Development Process Steps

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 (Universal Joint Task List Manual). 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 Appendix 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 CBA Handbook reference listed in Appendix B (References and Information Sources).

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), if identified, are being addressed?
- What are the desired effects?
- What objectives, major operations, or activities are intended 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 not uncommon 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. Appendix 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 that should be considered by the analyst when identifying attributes for a task.

Similar to tasks, attributes may not be independent. 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 have been identified, the analyst identifies the operational conditions associated with each task. Operational conditions can be described in terms of factors and descriptors (Universal Joint Task List Manual). 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 that is 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 that are 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 have been 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 Handbook reference listed in Appendix B 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, that are needed to sufficiently 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) is needed 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 that will be 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.

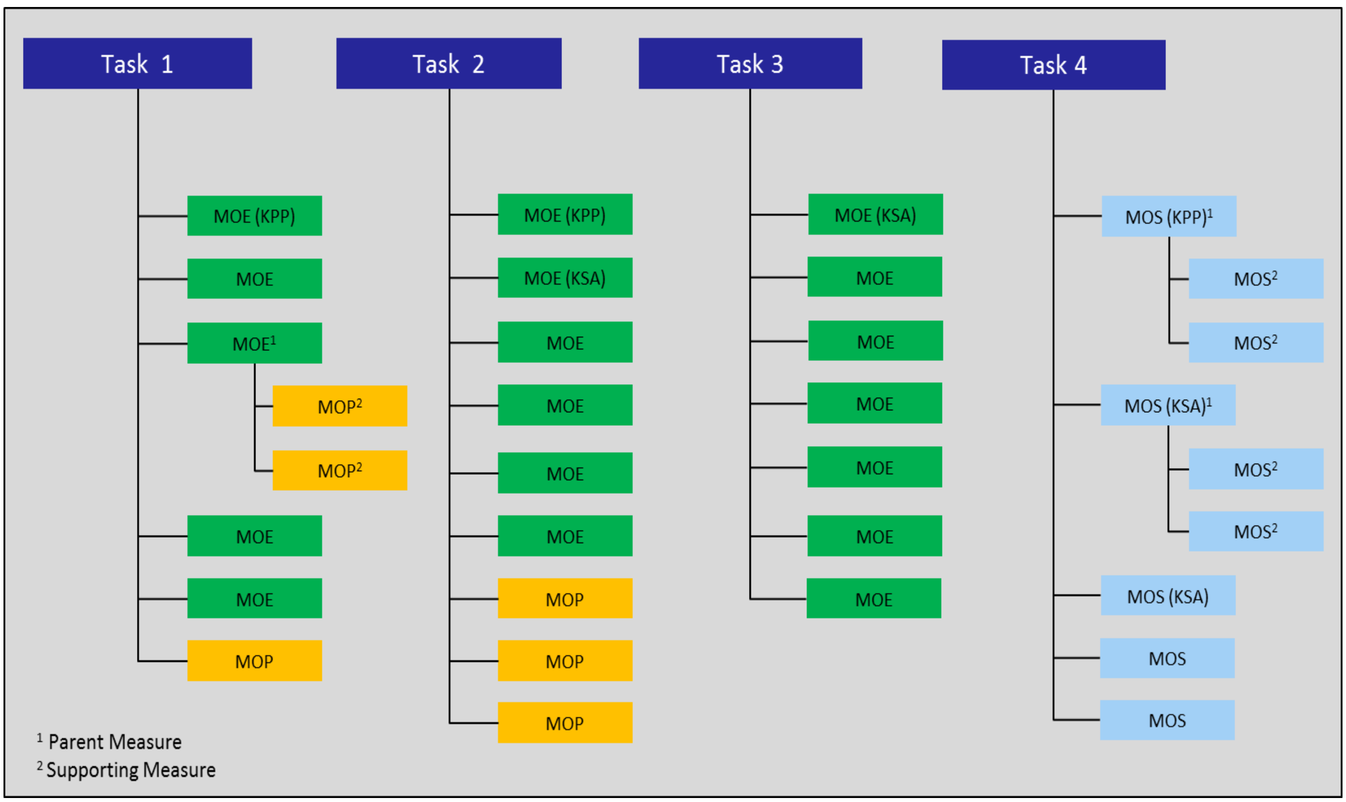


Figure 3-2: Measure Dendritic Example

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 Capability Production Document (CPD)). The analyst should review these documents, if developed, to identify user-established criteria that are relevant to the measures being assessed in the study.

When user-established criteria do not exist, criteria must be developed to enable the analyst to assess the measure. These criteria are referred to as identified criteria. Sources of information that can be used to develop identified criteria include CONOPS, CONEMPs, Tactics, Techniques, and Procedures (TTPs), and 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 all 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, 99^{th} 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	95^{th} 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	95^{th} 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	95^{th} 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 Measure Analysis and Rating

This chapter begins with an explanation of the measures framework and its key components. Next, there is a discussion of data collection and analysis methods the analyst can use to address measures. Finally, an approach to rating measures and guidelines for avoiding analysis pitfalls are presented.

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 <u>individual</u> 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. Appendix 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 Appendix F. Information about developing measurement scales for questions used in expert elicitation can be found in Appendix 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), and
- 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 Appendix B (References and Information Sources). Information about developing measurement scales for questions used in survey research can be found in Appendix 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.

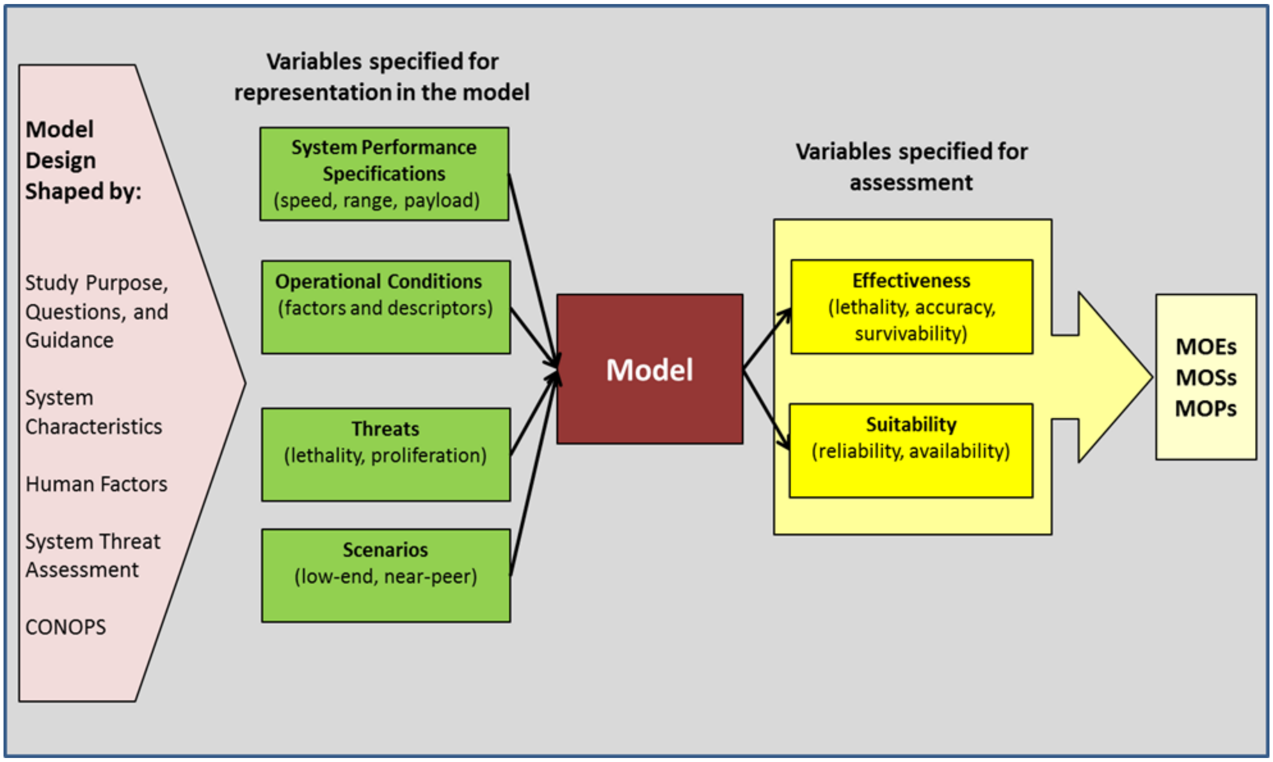


Figure 4-1: Model Variable Relationship

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 Appendix 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 AF/A9 Studies and Analyses, Assessments, and Lessons Learned website listed in Appendix 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 the CCTD document requirements, see the *Concept Characterization and Technical Description (CCTD) Guide* by SAF/AQ listed in Appendix B (References and Information Sources).

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.

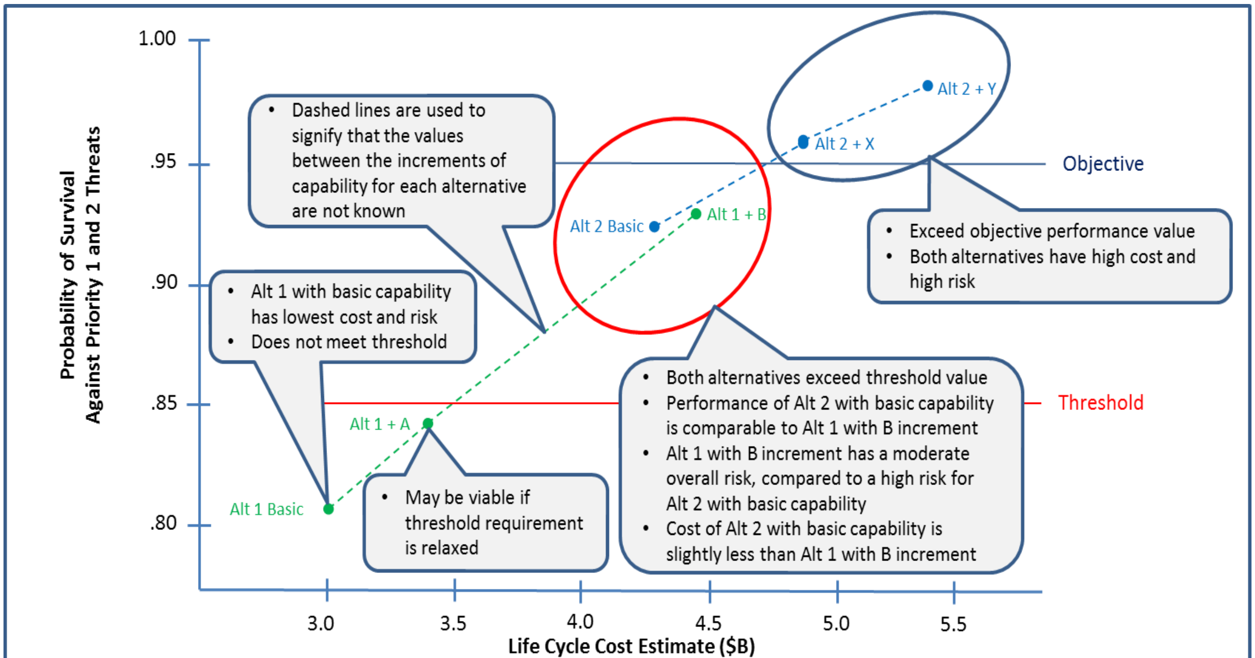
Another approach entails altering the operational conditions or scenarios to assess capabilities and limitations of systems in different environments. Using the results of the analysis, the analyst can determine how robust a system is in a wider range of operational conditions and scenarios. Whatever 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 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
    1. Criteria and criteria reference or rationale
    2. Measure rating
    3. Measure rating discussion
      a) Rationale or justification for rating
      b) 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.

Appendix 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
CPD	Capability Production Document
DAU	Defense Acquisition University
DOTmLPF-P	Doctrine, Organization, Training, materiel, Leadership and Education, Personnel, Facilities, and Policy (in this version of the acronym, "m" refers to existing materiel in the inventory)
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
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

Appendix B: References and Information Sources

Government Documents

AFH 33-337, *The Tongue and Quill*, 1 August 2004

AFOTECMAN 99-101, *Operational Test Processes and Procedures*, 6 August 2010

AFI 10-601, *Operational Capability Requirements Development*, 6 November 2013

AFI 10-604, *Capabilities-Based Planning*, 10 May 2006

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

AFI 61-101, *Management of Science and Technology*, 14 March 2013

Analysis of Alternatives Handbook: A Practical Guide to the Analysis of Alternatives, Office of Aerospace Studies, 10 June 2013

Capabilities-Based Assessment (CBA) Handbook: A Practical Guide to the Capabilities-Based Assessment, Office of Aerospace Studies, 10 March 2014

CJCSI 3170.01H, *Joint Capabilities Integration and Development System*, 10 January 2012

CJCSM 3500.04F, *Universal Joint Task List Manual*, 1 June 2011

Concept Characterization and Technical Description (CCTD) Guide, Assistant Secretary of the Air Force for Acquisition (SAF/AQ), 27 October 2010.

Defense Acquisition University (DAU) Glossary, Glossary of Defense Acquisition Acronyms and Terms, 13th Edition, November 2009.

DoDD 5000.01, *The Defense Acquisition System*, 8 December 2008

DoDI 5000.02, *Operation of the Defense Acquisition System*, 8 December 2008

Joint Publication (JP) 1-02, *Department of Defense Dictionary of Military and Associated Terms*, 8 November 2010

JS/J8, *Manual for the Operation of the Joint Capabilities Integration and Development System*, 19 January 2012 (JCIDS Manual)

JS/J8, *Manual for the Operation of the Joint Capabilities Integration and Development System*, 2014 near-final draft, (JCIDS Manual)

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

Web Links

AF/A5R-P Requirements: https://www.my.af.mil/gcss-af/afp40/USAF/ep/globalTab.do?command=org&channelPageId=-569424&pageId=681742

AF/A9 – Studies and Analyses, Assessments, and Lessons Learned: https://www.my.af.mil/gcss-af/USAF/ep/globalTab.do?channelPageId=s6925EC13500D0FB5E044080020E329A9

ACQ Notes – The JCIDS Process: http://www.acqnotes.com/Acquisitions/Capabilities%20Based%20Assessment%20(CBA).html

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

Air Force Portal: https://www.my.af.mil/

Defense Technical Information Center: www.dtic.mil

Defense Acquisition University: http://www.dau.mil/default.aspx

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

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.

Appendix 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 2012)

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)

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. (AFI 10-601)

Capability – the ability to complete a task or execute a course of action under specified conditions and level of performance. (draft JCIDS Manual 2014)

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. (draft JCIDS Manual 2014)

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 considered to be "draft" or "proposed" until validated by the appropriate authority. (JCIDS Manual 2012)

Conditions (Operational) – describes the environment under which the mission will be performed. (Universal Joint Task List Manual)

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). (AFOTECMAN 99-101)

Data – individual measurements that are used to compute the metric for a measure. (AFOTECMAN 99-101)

Key Performance Parameter – performance attributes of a system considered critical or essential to the development of an effective military capability. (JCIDS Manual 2012)

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 2012)

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. (AFOTECMAN 99-101)

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. (AFOTECMAN 99-101)

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. (AFI 10-601)

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 2012)

Stakeholder – any agency, Service, or organization with a vested interest (a stake) in the outcome of the analysis. (OAS AoA Handbook)

Task – describes what is expected to be performed and is commonly expressed as an action or activity. (Universal Joint Task List Manual)

Threshold – a minimum acceptable operational value of a system capability or characteristic below which the utility of the system becomes questionable. (AFI 10-601)

Appendix D: Operational Attributes by Joint Capability Area

Examples shown are from the draft JCIDS Manual 2014, Appendix A to Enclosure A.

Battlespace Awareness	Force Application	Building Partnerships	Protection
Accuracy	Adaptability	Agility	Effectiveness
Adaptability	Capacity	Breadth	Integration
Comprehensiveness	Flexibility	Depth	Networkability
Credibility	Mobility	Effect	Persistence
Innovativeness	Persistence	Flexibility	
Integration	Precision	Persistence	
Interoperability	Scalability	Utility	
Persistence	Security		
Survivability	Survivability		
Timeliness	Timeliness		

Command and Control	Logistics	
Accessibility	Accountability	Agility
Accuracy	Attainability	Capacity
Agility	Economy	Effectiveness
Completeness	Enduring	Expeditionary
Interoperability	Flexibility	Persistence
Operational Trust	Precision	Reliability
Relevance	Responsiveness	Simplicity
Robustness	Survivability	Sustainability
Security	Tailorability	Visibility
Simplicity	Velocity	
Timeliness		
Understanding		

Net Centric		
Accessibility	Accuracy	Agility
Availability	Capacity	Completeness
Controllability	Expeditionary	Flexibility
Integration	Interoperability	Latency
Maintainability	Reconfigurability	Relevance
Reliability	Responsiveness	Robustness
Scalability	Security	Survivability
Throughput	Timeliness	Visibility

Appendix E: Literature Review Sources of Information

1. Contract Studies Registry Program: https://reg.dtic.mil/login/generic.jsp?TYPE=100728833&REALMOID=06-0b4e5624-5088-103f-84828396ac570cb3&GUID=&SMAUTHREASON=0&METHOD=GET&SMAGENTNAME=olHaYdeyHv4zVMCSUwGWlUHU1xuFRpPZyP9UtkOzntAsnDhgeCHhtOcc0sihXLiM&TARGET=SMhttps%3a%2f%2fwww%2edtic%2emil%2f

2. Joint Lessons Learned Information System The Air Force-Joint Lessons Learned Information System: https://www.jllis.mil/usaf SIPRNET, http://www.jllis.smil.mil/usaf

3. DTIC: www.dtic.mil

4. Information and Resource Support System (IRSS): https://www.my.af.smil.mil/IRSS/irss7/pkg_portal.prc_main (requires SIPRNet Air Force Portal account, as well as permission from HAF/A5R)

5. Defense Acquisition University (ACQuipedia): https://dap.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=d5461b4c-2887-4be8-8cd9-b09920308670

6. Better Buying Power: http://bbp.dau.mil/

7. Rand Corp: www.rand.org

8. The Knowledge Management/Decision Support system (KM/DS): For instructions go to the JCIDS NIPRNet page: https://intellipedia.intelink.gov/my.policy

9. Manual for the Operation of the Joint Capabilities Integration and Development System, 19 January 2012 (JCIDS Manual) Enclosure A: https://intellipedia.intelink.gov/wiki/JCIDS_Manual#Latest_Approved_JCIDS_Documents

Appendix 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 appendix 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 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.

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 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 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 appendix 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.

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.

Appendix 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 appendix 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 appendix.

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.

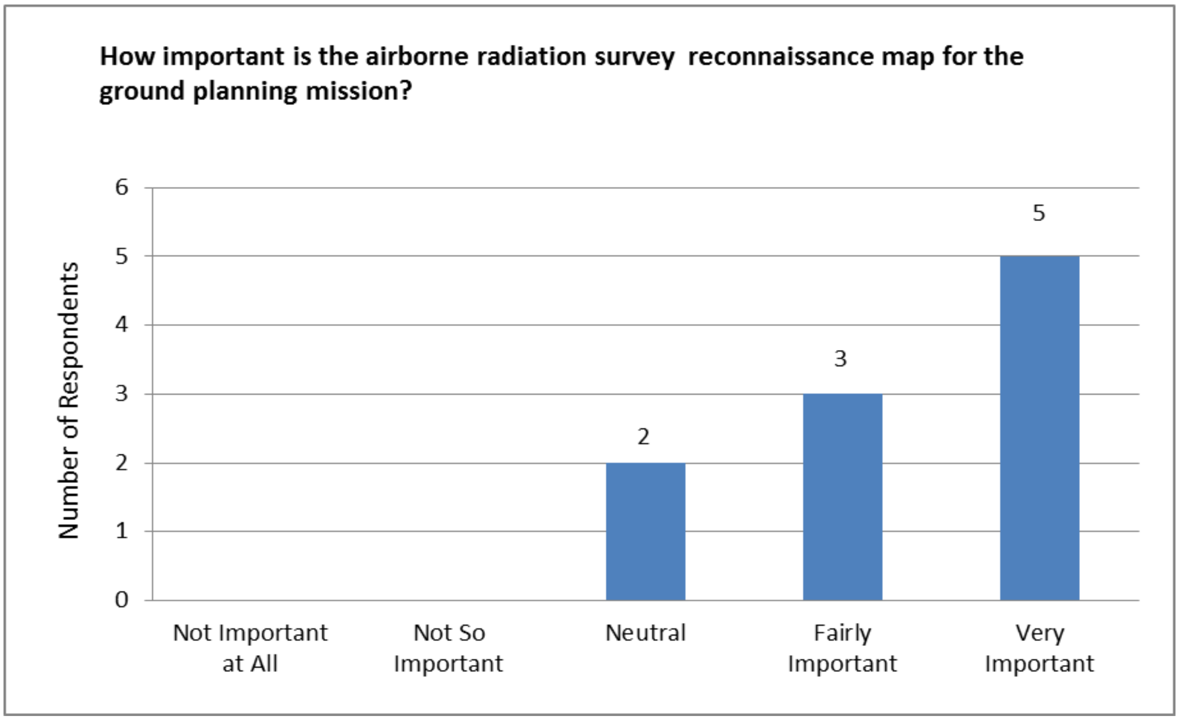


Figure G-1: Bar Chart Display without "Not Applicable" Selections

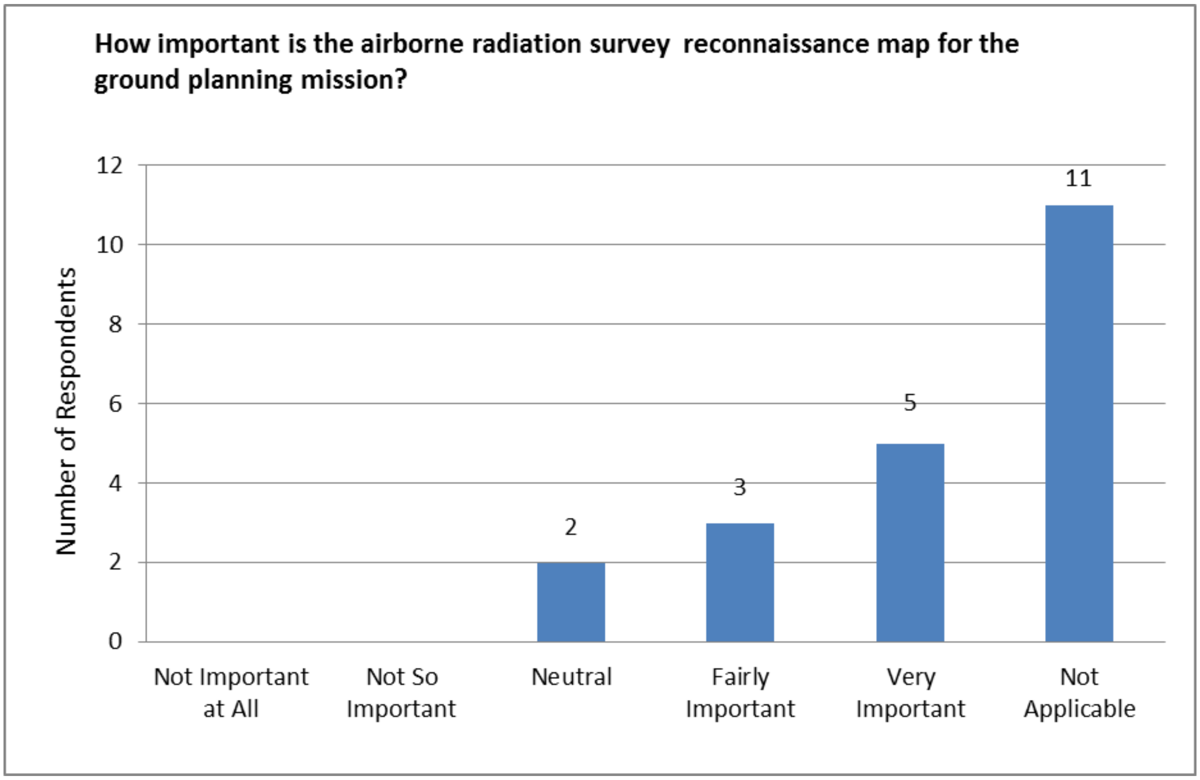


Figure G-2: Bar Chart Display with "Not Applicable" Selections

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.

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

This appendix 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 appendix.

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 use 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.

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.

THIS PAGE INTENTIONALLY LEFT BLANK