

**Standardized Technology Evaluation Process (STEP)
User's Guide and Methodology for Evaluation Teams**

Sarah Brown

May 2007

Table of Contents

1	Introduction	4
1.1	Purpose	4
1.2	Background	5
1.3	Intended Audience	5
1.4	How to Use this Document	5
2	STEP Methodology	7
2.1	Evaluation Phases	7
2.2	STEP Workflow	9
2.3	Tailoring STEP	10
2.3.1	STEP Workflow for Small Evaluation Teams	10
2.3.2	STEP Workflow for Single Product Evaluations	11
3	Guidance for Successful Evaluations	12
3.1	Methods Used to Evaluate and Score Products	12
3.1.1	Establishing Evaluation Criteria	13
3.1.2	Scoring the Products	14
3.1.3	Computing Weights	15
3.1.4	Computing the Overall Score for Each Product	19
3.2	Communication throughout the Evaluation Process	20
3.3	Ensuring Evaluation Integrity	20
3.4	Creating an Evaluation Timeline	21
4	Phase 1: Scoping and Test Strategy	22
4.1	Action: Conduct Preliminary Scoping	22
4.2	Action: Scoping with Government Sponsor	23
4.3	Action: Perform Market Survey/Tool Selection	24
4.4	Action: Determine Test Architecture	25

4.5	Action: Draft High-Level Test Plan	25
4.6	Check Point – Phase 1	26
5	Phase 2: Test Preparation	28
5.1	Action: Establish Evaluation Criteria, Priorities, and Test Procedures	28
5.2	Action: Perform Government Requirements’ Mapping	29
5.3	Action: Enhance and Finalize Test Plan	30
5.4	Action: Acquire Necessary Hardware and Software	31
5.5	Action: Hold Technical Exchange Meeting (TEM) (optional)	32
5.6	Check Point – Phase 2	33
6	Phase 3: Testing, Results, and Final Report	35
6.1	Action: Conduct Testing and Compile Results	35
6.2	Action: Perform Crosswalk	36
6.3	Action: Share Results with Vendors	36
6.4	Action: Deliver Final Report	37
6.5	Check Point – Phase 3	39
7	Acknowledgments	41
8	References	41
Appendix A Acronym and Definition List		43
Appendix B STEP Templates		44

1 Introduction

1.1 Purpose

MITRE conducts numerous technology evaluations for its sponsors each year, spanning a wide range of products and technologies. In order to keep pace with rapidly changing technology and sponsor needs, MITRE evaluation teams require a well-defined evaluation process that is efficient, repeatable, and as objective as possible.

The benefits of following a standardized, effective process include:

- Consistency and improved traceability through fixed steps and deliverables
- Improved efficiency leading to less effort required per evaluation
- Defensible, repeatable results
- Better communication within and among evaluation teams
- Evaluations that can be compared and shared more easily across the sponsor base
- An opportunity to develop guidance and document lessons-learned for future evaluations

The Standard Technical Evaluation Process (STEP) developed in G024 outlines a rigorous process for technology evaluations of one or more COTS products¹. It applies to a variety of areas of technology and provides substantial benefits for evaluation teams and their government sponsors.

STEP aims to provide:

- A process that can be used in a broad range of technology evaluations
- Standard deliverables to achieve consistency, traceability, and defensibility of the evaluation results
- Guidelines to assist teams in developing goals, documenting findings, and addressing challenges
- A process that is recognized as comprehensive and fair

This document presents STEP and offers a guide to evaluation teams who wish to use it. From preliminary scoping to eventual integration and deployment, STEP guides teams in producing high quality reports, thorough evaluations, and defensible results.

¹ *Technology evaluation* is used in this document to refer to evaluations of multiple products providing the same capability. *Product evaluation* is used to refer to an evaluation of a single product.

1.2 Background

In 2004, the MITRE Intelligence Community Test and Integration Center in G024 began developing STEP in an effort to track past evaluation work and ensure quality, objectivity, and consistency in future evaluations. Since that time, STEP has been used successfully in G024 as well as in G025, G027, and G151 evaluation tasks.

The four phases of STEP follow a common framework for conducting a technology evaluation. In developing and refining STEP, a variety of resources and subject matter experts were sought (see references in Section 8) within and outside of MITRE to gain a broader understanding of evaluation theory and practice. The STEP workflow and methodology incorporate many of these practices and recommendations.

1.3 Intended Audience

This document is intended for MITRE project leads and engineers conducting technology evaluations of one or more products and is suitable for experienced as well as first-time evaluators. Although STEP was designed originally for several G024 security tool evaluations, the process and methodology is applicable to any software or information technology evaluation. Because evaluations may vary significantly in size and scope, STEP presents options for evaluation teams that would like to work in parallel for improved efficiency, as well as for smaller teams that wish to work together through each stage. Together, the STEP workflow and methodology provide a comprehensive resource for teams wishing to standardize their evaluations and structure their daily activities.

1.4 How to Use this Document

Section 2 of this document provides guidance on four major challenges in technology evaluations: using an established scoring method, communicating with the sponsor, ensuring integrity and defensibility, and forming a realistic evaluation timeline.

The remainder of the document provides specific information for executing each STEP action. The presentation in this document is based on the CEM Project Leader Handbook [8]. There is a chapter for the three main STEP phases and the chapters are designed so that the reader can quickly locate information about a specific action. Each chapter contains:

- An overview of the phase
- A section for each action within the phase
- For each action:
 - **Description:** A description of the action and specific work to complete
 - **Lessons-learned:** Guidance for successfully completing the action
 - **Templates and Sample Deliverables:** A list of templates and deliverables from past evaluations to assist teams in documenting their work

The final STEP phase, Phase 4: Integration and Deployment, is outside the scope of this document and is not addressed in detail. Phase 4 applies if an evaluation results in a purchase decision by the sponsor. In this case, the sponsor determines the specific actions required.

2 STEP Methodology

2.1 Evaluation Phases

The STEP process defines evaluations according to three main phases: (1) Scoping and Test Strategy, (2) Test Preparation, (3) Testing, Results, and Final Report, and a fourth, optional phase (4) Integration and Deployment that is determined by the sponsor on a case-by-case basis (Figure 1). Each STEP phase has different objectives, actions and associated document deliverables.

Checkpoints, or control gates, separate the phases, and each phase must be completed before the next one is begun. These control gates help to ensure evaluation integrity. For instance, teams must establish their evaluation criteria and test strategy (Phase 2) before installing or testing the evaluation products (Phase 3). It is critical that the team solidify their evaluation criteria before starting hands-on product testing. This avoids the potential for introducing bias into the evaluation criteria based on prior knowledge of a given product's features or design.

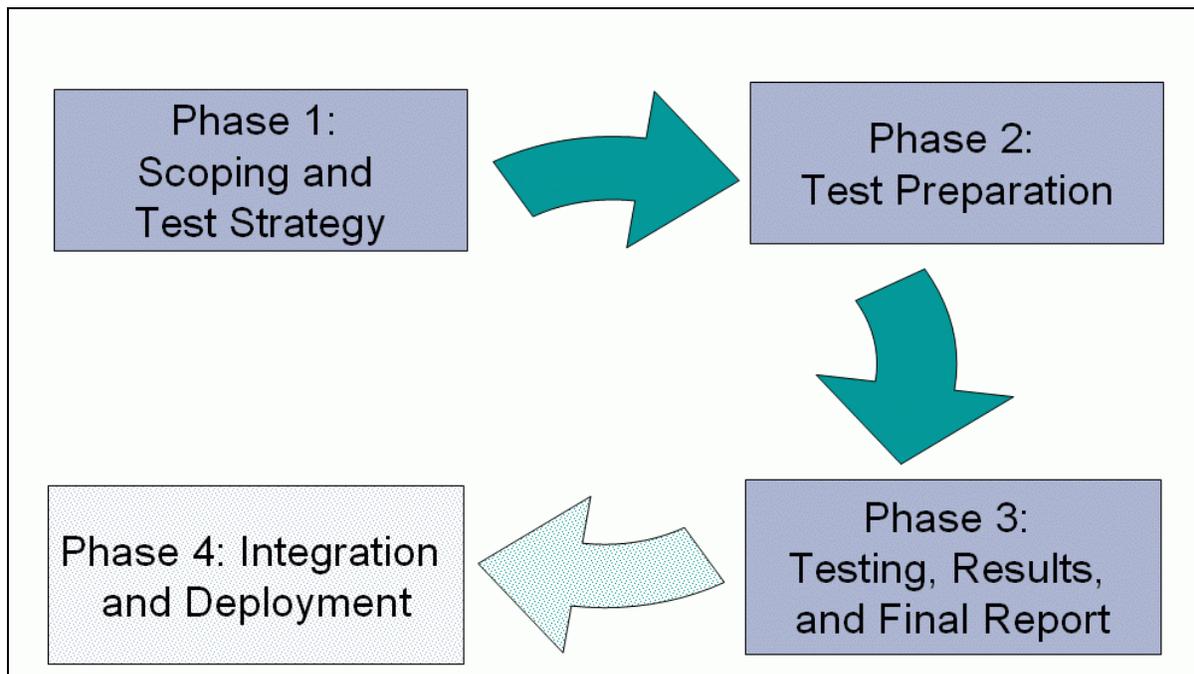


Figure 1: Four Phases of STEP

Below are short descriptions of each phase:

1. **Scoping and Test Strategy.** During this phase, the evaluation team gains an understanding of the mission objectives and technology space, and settles on key requirements through scoping with the government sponsor. The team produces a project summary to help clarify the objectives and scope, and performs a market survey to identify potential products in the technology area. The evaluation team

works with the sponsor to select a list of products for further evaluation based on the market survey results, evaluation timeline, and resources available. To prepare for testing, the team produces a project summary and high-level test plan.

2. **Test Preparation.** After selecting the products to evaluate and obtaining concurrence from the sponsor, the evaluation team works to acquire the evaluation products from the vendors, and any additional infrastructure that is required for testing. This includes signing non-disclosure agreements (NDAs), establishing vendor points of contact, and meeting with the vendor to discuss the test plan. At the same time, the team develops a full set of evaluation criteria that the products will be tested against and any scenario tests² that will be performed. The evaluation team then installs the products in the test environment, and engages the vendor as technical questions arise. The team may wish to hold a technical exchange meeting (TEM) to gain further insight and background from subject matter experts.
3. **Testing, Results, and Final Report.** In this phase, the evaluation team tests and scores the products against all of the test criteria. The team must ensure that testing for each product is performed under identical conditions, and must complete a full crosswalk of the scores for each product requirement after testing to ensure scoring consistency. Following the crosswalk, evaluation team members conduct individual meetings with each vendor to review their findings, correct any misunderstandings about their product's functionality, and retest if necessary. The team produces a final report that incorporates the evaluation results and any supporting information.
4. **Integration and Deployment**³. The final evaluation report submitted to the government provides a data source to assist in decision-making, but is not a proposal to purchase specific products. If the government decides to purchase a product, the evaluation team works with the government and other commercial contractors to assist in deploying and integrating the solution into the operational environment. Actions in this phase may include developing configuration guidance and supporting documentation. .

² In a scenario test, product performance is determined in situation that models a real-world application. The evaluation team must ensure that each product tested receives the same data and is in the same environment. Test results will be repeatable only to the extent that the modeled scenario and data can be reproduced.

³ Phase 4 is outside the scope of this document. It is not addressed in later chapters.

2.2 STEP Workflow

Figure 2 presents the full STEP workflow. STEP is comprised of four phases separated by checkpoints. Within each phase, most actions can be completed in parallel so that teams can maximize their efficiency. The highlighted actions result in major document deliverables for the sponsor. Appendix A of this guide contains templates for completing each STEP action.

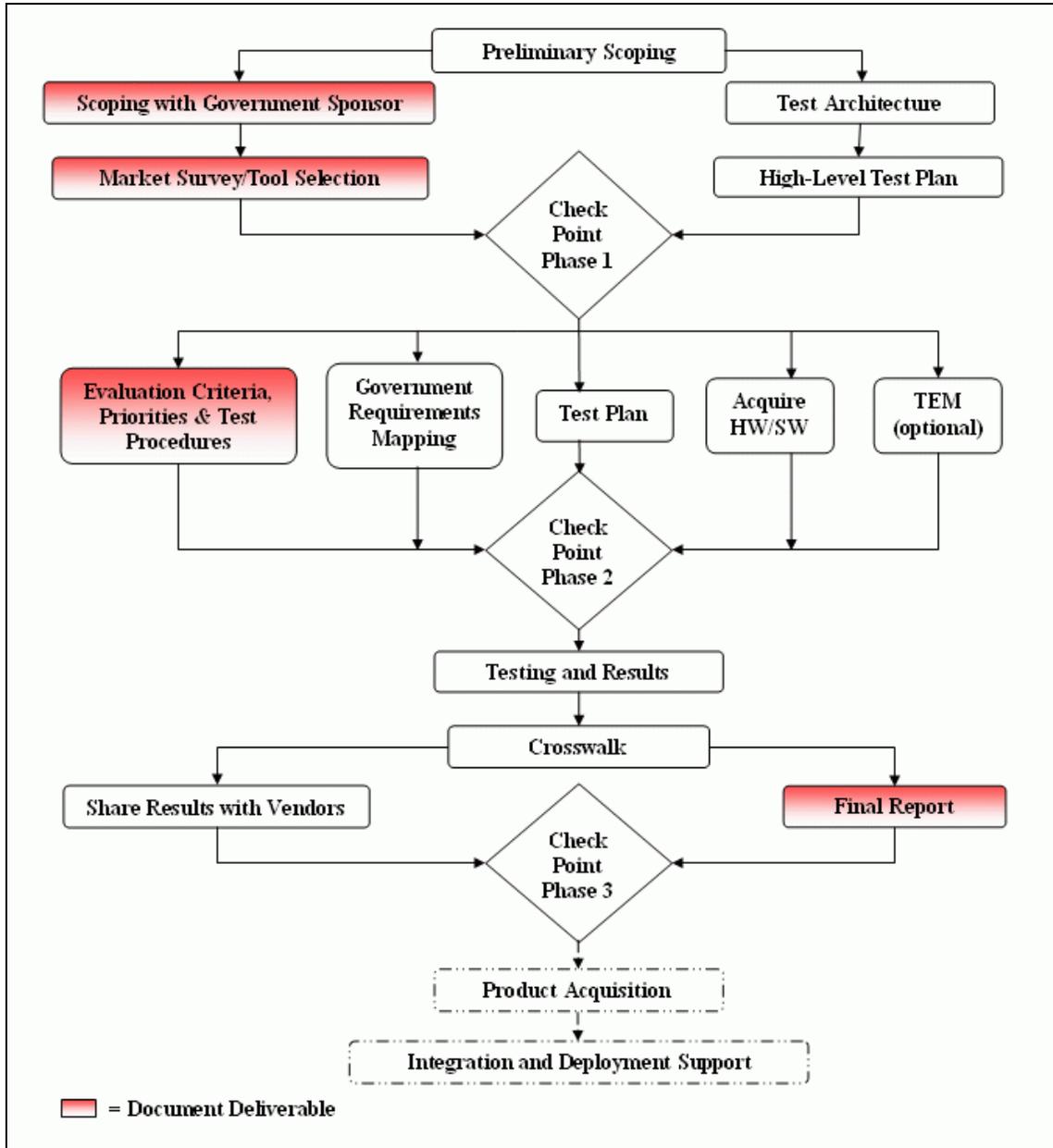


Figure 2: Full STEP Workflow

2.3 Tailoring STEP

2.3.1 STEP Workflow for Small Evaluation Teams

For small evaluation teams that wish to perform the STEP actions in a linear order, Table 1 presents a recommended workflow.

Table 1: Recommended Linear STEP Workflow

STEP Phase	Section	Action
Phase 1 - Scoping and Test Strategy	§ 4.1	Conduct Preliminary Scoping
	§ 4.2	Scoping with Government Sponsor
	§ 4.3	Perform Market Survey/Tool Selection
	§ 4.4	Determine Test Architecture
	§ 4.5	Draft High-Level Test Plan
	§ 4.6	Check Point – Phase 1
Phase 2 - Test Preparation	§ 5.1	Establish Evaluation Criteria, Priorities & Test Procedures
	§ 5.2	Perform Government Requirements’ Mapping
	§ 5.3	Enhance and Finalize Test Plan
	§ 5.4	Acquire Necessary Hardware and Software
	§ 5.5	Hold Technical Exchange Meeting (TEM) (optional)
	§ 5.6	Check Point – Phase 2
Phase 3 - Testing, Results, and Final Report	§ 6.1	Conduct Testing and Compile Results
	§ 6.2	Perform Crosswalk
	§ 6.3	Share Results with Vendors
	§ 6.4	Deliver Final Report
	§ 6.5	Check Point – Phase 3
Phase 4 - Integration and Deployment	none	Determined by sponsor

2.3.2 STEP Workflow for Single Product Evaluations

While the full STEP workflow is designed for technology evaluations (evaluations involving multiple products), it can be modified for teams performing a single product evaluation. In this situation, Figure 3 provides a tailored workflow.

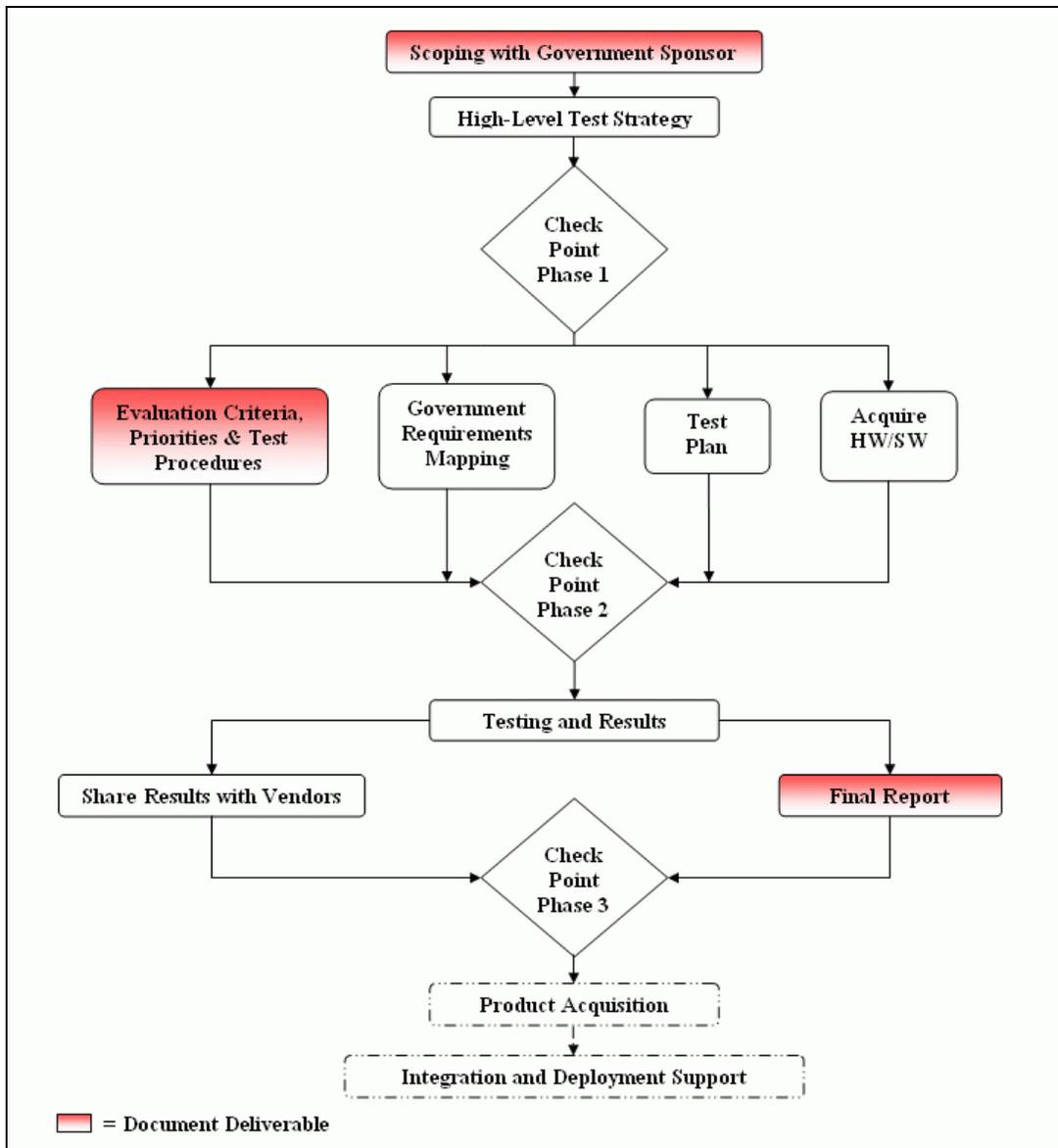


Figure 3: STEP Workflow for Single Product Evaluations

3 Guidance for Successful Evaluations

In developing STEP, project leads identified several key challenges in conducting technology evaluations. The following subsections address the four challenges identified by MITRE evaluation teams that are critical to ensuring an evaluation's success:

- Methods used to evaluate and score products,
- Communication during the evaluation process,
- Ensuring evaluation integrity, and
- Creating an evaluation timeline.

These challenges were echoed and addressed in several literature searches on decision making. As stated in an article [6] on methods and best practices in evaluating alternatives:

“There are many potential mistakes that can lead one awry in a task...Some concern understanding the task. Others concern structuring the decision problem to be addressed. Still others occur in determining the judgments necessary to specify the [scores]... These mistakes frequently cascade... ‘When this occurs, the [scores] provide little or no insight, contribute to a poor decision, and result in frustration with the decision process.’”

3.1 Methods Used to Evaluate and Score Products

In a technology evaluation, teams must evaluate and score products against a set of evaluation criteria in order to determine the best choice to meet their sponsor's needs. Teams must produce a clear assessment of the products and provide a rationale that can be used to make and justify decisions. The process involves

1. establishing a set of evaluation criteria and, as appropriate, dividing the criteria among a set of categories,
2. determining a scheme for scoring products against the evaluation criteria
3. providing a set of numerical weights to determine the relative importance of the criteria and evaluation categories
4. computing the overall score for each product

Teams often use a spreadsheet such as the one in Table 2 to track the evaluation criteria, scores, and weights, and calculate the total weighted scores for each product (see Appendix B for this Evaluation Criteria Template).

Table 2: Spreadsheet for capturing evaluation criteria, weights, and scores

#	Evaluation Criteria	Description of How to Test the Criteria	Weight	<product P1 scores>	<product P2 scores>	<product P3 scores>	<product P4 scores>	<product P5 scores>
1.0	Category 1 Title							
1.1	Criteria A	-description-						
1.2	Criteria B	-description-						
1.3	Criteria C	-description-						
1.4	Criteria D	-description-						

The following subsections provide guidance for accomplishing steps 1- 4 above. This guidance comes from the multi-attribute utility (MAU) analysis, within the mathematical field of decision analysis. Decision analysis is concerned with providing a mathematical framework for decision making, so that decision makers can rigorously and consistently express their preferences, in such a way that their results can be readily and logically explained.

Multi-attribute utility (MAU) analysis [1, 2, 3, 4, 5, 6, 7, 10, and 14] is a well-established decision analysis method that specifically addresses how to select one alternative from a set of alternatives, which is akin to selecting a particular product from a set of products in a given technology area. MAU analysis follows steps 1- 4 above to compute the overall score, or utility, of each alternative under consideration. By following the rules and principles of MAU analysis, evaluation teams can perform straightforward, rigorous, and consistent decision making. Furthermore, teams can back up the integrity of their results through an established scoring method that is recognized as comprehensive and fair.

3.1.1 Establishing Evaluation Criteria

In preparing for the evaluation testing, the first step is to establish the evaluation criteria. This is a key step, because at the end of the evaluation, the results will be a reflection of how well the team created their evaluation criteria. In order to generate these criteria, the team should conduct independent research and request guidance on all aspects and objectives of the problem from the government sponsor and subject matter experts. Through this research, the team will ensure that the sponsor’s primary needs/wants are addressed, as well as critical functional (e.g. security) capabilities or nonfunctional (e.g., policy, vendor support) issues.

Evaluation criteria should be specific, Boolean (two-valued) types of questions that are clearly stated and can be clearly tested. The following tips are provided for writing individual criteria statements. First, use the “who shall what” standard form to prevent misunderstanding. In other words,

The
(service provider or product)
shall
(do, perform, provide, weigh, or other verb,
followed by a description of what).

Figure 4: Standard form for writing the evaluation criteria

In writing these statements, avoid the following pitfalls listed in [13]:

- Ambiguity – write as clearly as possible so as to provide a single meaning
- Multiple criteria – criteria that contain conjunctions (and, or, with, also) can often be split into independent criteria
- Mixing evaluation areas – do not mix design, system, user, and vendor support criteria in the same evaluation category.
- Wishful thinking – “Totally safe”, “Runs on all platforms”.
- Vague terms – “User friendly”, speculative words such as “generally”, “usually”

In addition to the evaluation criteria statements, provide a description of how each criterion will be tested. Following these tips will help ensure that each evaluation criterion is carefully written, independent, and clearly states *what* is tested, *how* it is tested, and the desired outcome.

3.1.2 Scoring the Products

The next step is to determine how products will be scored against the evaluation criteria. For example, teams could use the following function u_i :

- $u_i(a_i) = 0$ if a product does not meet evaluation criteria a_i
- $u_i(a_i) = 1$ if a product partially meets evaluation criteria a_i
- $u_i(a_i) = 2$ if a product fully meets evaluation criteria a_i

This function is a constructed scale because each point is explicitly defined. Constructed scales are often useful because they allow both quantitative and qualitative criteria to be measured. Teams may prefer to assign scores based on a standard unit of measure (e.g., time, dollars), a complex function, or another function type.

By convention, in MAU analysis, any scoring function should be normalized so that the scores fall in the range from 0 to 1. Normalizing the above constructed scale gives:

- $u_i(a_i) = 0$ if a product does not meet evaluation criteria a_i
- $u_i(a_i) = .5$ if a product partially meets evaluation criteria a_i
- $u_i(a_i) = 1$ if a product fully meets evaluation criteria a_i

Therefore, in the above example, a product that fully meets a criterion during testing will receive a score of 1, a product that partially meets a criterion will receive a score of .5, and a product that does not meet a criterion will receive a 0 for that item. These are not the only possible scale values. In this case we have a discrete set of three values. We could have a larger discrete set or a continuous set between 0 and 1.

3.1.3 Computing Weights

The final step is to assign weights w_i to each criterion. These weights serve as scaling factors to specify the relative importance of each criterion. Because they are scaling factors that specify relative importance in the overall set of criteria, they should be nonnegative numbers that sum to 1.

There is no “best” method for choosing weights. The choice depends on the principles and axioms that the decision maker wishes to follow, level of detail desired for the weights, and the computing resources available for calculating the weights.

A variety of methods have been proposed for eliciting weights [1, 2, 3, 4, 10, and 14]. These methods include:

- Weighted Ranking
- Analytic Hierarchy Process (AHP)
- Trade-off method (also called Pricing Out)
- Paired Comparison (also called Balance Beam method)
- Reference Comparison

These methods are compared in Figure 5 below and the Paired Comparison and Reference Comparison methods are recommended for use by MITRE evaluation teams.

The first three methods, weighted ranking, AHP, and the trade-off method, are not recommended in this guide for the following reasons. Both weighted ranking [2, 9] and AHP [5, 10] are popular methods, but they can be manipulated in ways that result in certain basic logical flaws, and as a result, are often rejected by decision analysts as acceptable methods for computing weights [2, 4, 11, 14]. The Trade-Off method [2, 3, 6] is also a well-accepted method, but is not recommended because of the computational resources required to derive weights for more than 10 alternatives. Several commercial decision software packages are available that implement this method.

The Paired Comparison and Reference Comparison [3, 9, and 14] are recommended in this guide for use by evaluation teams because they are widely accepted and practical to perform by hand. The Paired Comparison is a good choice when deriving weights for 10-100 alternatives. Alternatively, the Reference Comparison method is a good choice when deriving weights for 100+ evaluation criteria. It requires fewer computations than Paired Comparison; however it provides less granular weights.

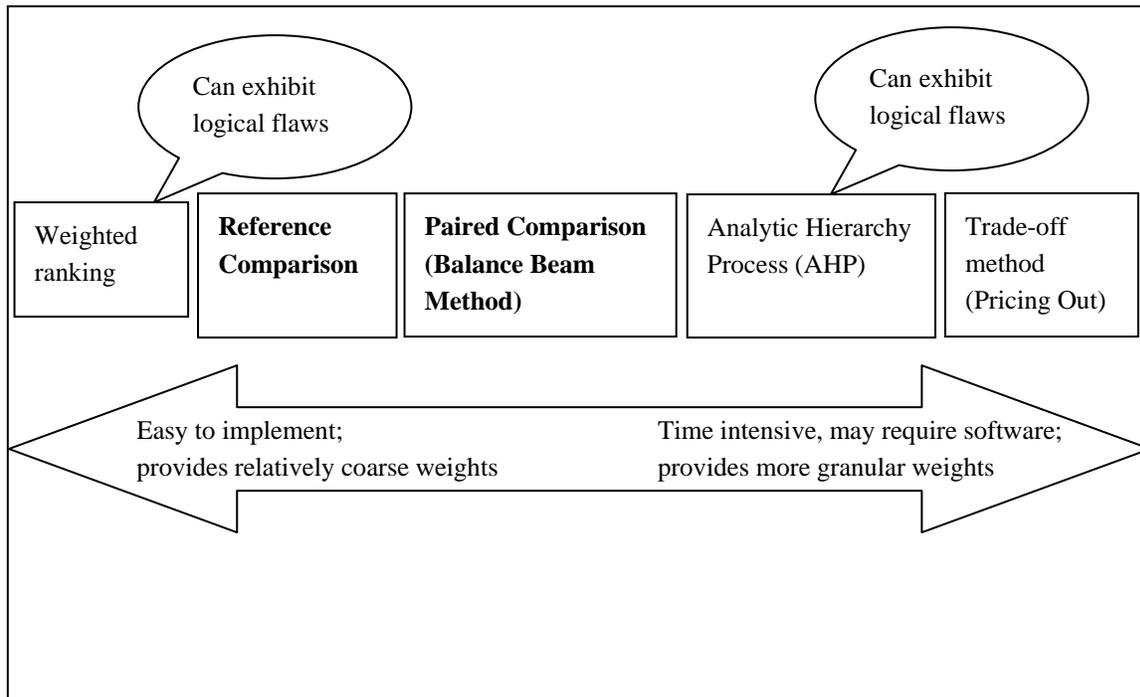


Figure 5: Comparison of Weight Assessment Methods. Reference Comparison and Paired Comparison are recommended in this Guide for evaluation teams

Paired Comparison:

This method is a good choice for deriving weights for 10-100 alternatives and is best explained with an example. Given a set of evaluation categories or a small set of evaluation criteria, determine a basic ordering from highest importance to least importance. Throughout these weight assessment methods, basic orderings and relative importance is decided by the team and will be subjective.

Example:

Most important = A

B

C

D

E

F

Least important = G

For example, in an evaluation of a security product, security is the most important category, followed by auditing, administration/management, and then vendor resources.

Starting with the alternative of highest importance, express its importance with the alternatives of lower importance in terms of a $<$, $=$, or $>$ relationship. There is no rule about coming up with this expression, it is determined by the evaluation team. Obtain an equality ($=$) relationship whenever possible to make it easier to solve the equations at the end. Repeat this with the alternative of next highest importance, until each alternative is expressed in terms of lower-order alternatives, as shown:

Paired Comparisons (Balance Beam Comparisons)	Relationship
$A < B + C$ $A = B + D$	$A = B + D$
$B > C + D$ $B < C + D + E$ $B < C + D + F$ $B = C + D + G$	$B = C + D + G$
$C < D + E$ $C < D + F$ $C > D + G$	$C > D + G$ $C < D + F$
$D = E$	$D = E$
$E > F + G$ $E = 1.5(F + G)$	$E = 1.5(F + G)$
$F = 2G$	$F = 2G$

Next, assign the lowest-order alternative (in this case, G) a value of 1. Then back solve the system of equations to determine values for the set of alternatives. The result in this example is:

$$A = 17.5$$

$$B = 11.5$$

$$C > 5.5 \text{ and } C < 6.5$$

$$D = 4.5$$

$$E = 4.5$$

$$F = 2$$

$$G = 1$$

Since the value for C is not exact, it can be approximated and assigned a weight of 6.

The sum of these weights is 47, so to normalize the values, divide each one by 47. The resulting numbers sum to 1 and give the weights. From A to G they are: 0.372, 0.245, 0.128, 0.096, 0.096, 0.043, and 0.020.

The paired comparison method can be used to find weights for the individual evaluation criteria and/or for the evaluation categories themselves. The table below shows the weights corresponding to individual evaluation criteria.

Table 3: Paired Comparison Weights shown on Evaluation Criteria Template

#	Evaluation Criteria	Description of How to Test the Criteria	Weight	<product P1 name>	<product P2 name>	<product P3 name>	<product P4 name>	<product P5 name>
1.0	Category 1							
1.1	Criteria A	-description-	0.372	0	0	0	0	0
1.2	Criteria B	-description-	0.245	0	0	0	0	0
1.3	Criteria C	-description-	0.128	0	0	0	0	0
1.4	Criteria D	-description-	0.096	0	0	0	0	0

Reference Comparison:

The Reference Comparison method is an alternative to the Paired Comparison and is a good alternative when calculating weights for 100+ criteria. Given a set of evaluation criteria, choose the evaluation criterion that is most important or significant in the set. Assign this criterion a value of 3. Using this as a reference, rank the remaining criteria as follows⁴:

- 3 = the criterion is as important as the “reference criterion”
- 2 = the criterion is slightly less important as the “reference criterion”
- 1 = the criterion is much less important than the “reference criterion”

Then, normalize these values so that they sum to 1.

For example, suppose values are assigned as follows:

$$A = 3$$

$$B = 3$$

⁴ It is not necessary to use the range from 1 to 3. The range can be less constrained or more constrained as needed.

C = 2

D = 2

E = 3

F = 1

G = 2

The sum of these weights is 16, so to normalize the values, divide each one by 16. The resulting numbers sum to 1 and give the weights. From A to G they are: 0.1875, 0.1875, 0.125, 0.125, 0.1875, 0.0625, and 0.125.

The reference comparison method can be used to elicit weights for the individual evaluation criteria and/or for the evaluation categories themselves. The table below shows the weights corresponding to individual evaluation criteria.

Table 4: Reference Comparison Weights on Evaluation Criteria Template

#	Evaluation Criteria	Description of How to Test the Criteria	Weight	<product P1 name>	<product P2 name>	<product P3 name>	<product P4 name>	<product P5 name>
1.0	Category 1							
1.1	Criteria A	-description-	0.1875	0	0	0	0	0
1.2	Criteria B	-description-	0.1875	0	0	0	0	0
1.3	Criteria C	-description-	0.125	0	0	0	0	0
1.4	Criteria D	-description-	0.125	0	0	0	0	0

3.1.4 Computing the Overall Score for Each Product

Once the evaluation criteria, product scores, and evaluation weights have been determined, the n th additive utility function is used to compute the overall score of each product, where n is the number of evaluation criteria.

As an example, the additive utility function with two evaluation criteria, a_1 and a_2 , is:

$$u(a_1, a_2) = w_1 u_1(a_1) + w_2 u_2(a_2)$$

The variables in the function are:

- u , represents the overall score of a product over two evaluation criteria, a_1 and a_2
- u_1 and u_2 , scoring function(s) for criteria a_1 and a_2 , respectively. For simplicity, teams can use the same scoring function for each criterion. The scoring function example from Section 3.1.2 demonstrated a constructed scale.

- w_1 and w_2 , individual weights assigned to each criterion by a weight assessment method. The process of eliciting weights was described in Section 3.1.3.

Therefore in summary, MAU analysis provides evaluation teams with a consistent, fairly rigorous approach for scoring products in a technology evaluation. Teams must establish the evaluation criteria; determine a scheme for scoring products; and weight the relative importance of each evaluation criterion and category. The results are the collective efforts of evaluation teams, and are therefore likely to have some inter-subjective consistency. After each product has been evaluated and scored, the n th additive utility function gives the overall score (or utility) for each product and an overall product ranking.

3.2 Communication throughout the Evaluation Process

A successful evaluation requires effective communication between the evaluation team and the sponsor, stakeholders, subject matter experts, and vendors throughout the evaluation process. The team must understand what the problem is and what the solution is intended to accomplish.

During each phase, evaluation teams should conduct status updates with the sponsor and stakeholders and/or subject matter experts, either in writing or as a briefing, to discuss and solicit feedback on the following items:

- Evaluation goals and objectives
- Initial product assessments
- Additional products or currently deployed solutions within the sponsor's environment worth considering
- Considerations/requirements for the sponsor's environment
- Evaluation criteria and the test plan

In order to facilitate consistent, well-presented work during an evaluation that is recorded for later reference, Appendix B provides STEP briefing and document deliverable templates for each phase of the evaluation. In addition to ensuring good communication throughout the evaluation, the STEP templates also assist the team in drafting their final report.

3.3 Ensuring Evaluation Integrity

It is critical that MITRE teams perform evaluations that are recognized as comprehensive and fair. A fundamental requirement to achieving evaluation integrity is consistent documentation of test data and methodology for review by the sponsor, stakeholders, and vendors if questions arise. The STEP actions and tips (Chapters 4-6) provide guidance for ensuring evaluation integrity. These guidelines include:

- Verifying all product information for a Market Survey/Tool Selection with product vendors, and requesting written explanations (by email) as needed

- Following the rules and principles for establishing evaluation criteria, scoring products, and weighting criteria, as explained in Section 3.1
- Finalizing evaluation criteria, including associated weights, test procedures, and expected outcomes/guidelines for scoring before testing is begun.
- Highlighting product strengths and weaknesses as they are indicated in the overall evaluation scores. That is, the evaluation team must be careful not to call out product strengths and weaknesses arbitrarily in the final report without quantitative results and/or justification to back up the claims.
- Documenting the evaluation using STEP templates for consistency

3.4 Creating an Evaluation Timeline

Scheduling is an important part of the evaluation process in order to establish realistic timelines and expectations. The STEP workflow allows teams to identify the individual actions and estimate the time required to complete each one. Teams may wish to break larger actions into smaller segments to ensure that all of the evaluation work is well defined [13]. Teams must also work with their sponsor to determine the appropriate number of products to be tested with the time and resources available. Successful planning and timelines throughout the project will result in managing the work required for the evaluation.

4 Phase 1: Scoping and Test Strategy

During this phase, the evaluation team gains an understanding of the mission objectives and technology space, and settles on key requirements through scoping with the government sponsor. The team produces a project summary to help clarify the objectives and scope, and performs a market survey to identify potential products in the technology area. The evaluation team works with the sponsor to select a list of products for further evaluation based on the market survey results, evaluation timeline, and resources available. To prepare for testing, the team produces a project summary and high-level test plan.

4.1 Action: Conduct Preliminary Scoping

Description and Activities	<p>Gather resources and information on the technology area to learn about the different vendors, products, and solutions currently available.</p> <p>Collect information from multiple sources, including:</p> <ul style="list-style-type: none">• Online research• Trade journals (Gartner, Burton, etc.)• Business reports (e.g., IDC and Gartner)• Sponsor meetings• Recommendations from subject matter experts (e.g., MITRE expertise search, MITRE Infosec list) <p>Distribute the initial list of products across the team for individual team members to research. Complete Preliminary Scoping sheets for each product to obtain company background and high-level information for each product.</p>
Lessons Learned	<p>Save all web pages, correspondence, and articles collected during the preliminary scoping action (along with dates and any other bibliographic information) as evidence of how each stage of the evaluation is performed. This information may also be included as references in the final report.</p> <p>Allow sufficient time for team members to become familiar with each product at a high level.</p> <p>It is not necessary to contact vendors at this point, unless the team cannot find basic information about the product from market literature.</p>

Templates and Sample Deliverables (Samples available upon request)	Preliminary Scoping template
--	------------------------------

4.2 Action: Scoping with Government Sponsor

Description and Activities	<p>Create a list of important scoping questions and/or base assumptions for the sponsor. These are key requirements or needs that the team feels should be addressed in the evaluation.</p> <p>Use disqualifiers and clear thresholds (i.e. go/no go requirements) to eliminate unsuitable options.</p> <p>Review these items with the sponsor and obtain any guidance and/or corrections. Build this list from information collected during the Preliminary Scoping.</p> <p>Discuss the priorities for the key criteria (e.g., “required”, “highly desired”). Document this information in the Scoping Questions with Sponsor template.</p>
Lessons Learned	<p>(Team lead) Present preliminary team findings and information to the sponsor and clarify your understanding of what evaluation is to be performed (scope and objectives), when the work is to be accomplished (project timeline), and who will be involved in the work (evaluation team members, team lead, sponsor, any additional stakeholders or standing boards).</p> <p>The purpose of this action is to review the teams’ preliminary findings and make sure that the sponsor and team are on the same page. The list of scoping questions may be greater or fewer than 10 items; this is an approximate size.</p> <p>This discussion should confirm the sponsor’s concerns and requirements to the team.</p> <p>Write statements of the key scoping criteria that can translate into evaluation criteria (completed in Phase 2).</p>

Templates and Sample Deliverables (Samples available upon request)	Scoping with Sponsor template Sample: Network IDS/IPS System Scoping Questions document
---	--

4.3 Action: Perform Market Survey/Tool Selection

Description and Activities	<p>After establishing the key scoping criteria with the sponsor (Action 2), use the Market Survey/Tool Selection template to create a matrix of initial products under consideration vs. key scoping criteria developed with the sponsor.</p> <p>Scoping criteria that the sponsor considers to be firm requirements should be assigned the highest weight/priority. Then, score each product according to the established scoring methodology presented in Section 3. The Paired Comparison method is recommended for determining weights.</p> <p>Contact product vendors to verify the information used to calculate the scores and save this correspondence for future reference should questions arise.</p> <p>Use the results of the Market Survey/Tool Selection to recommend a list of products for full evaluation.</p>
Lessons Learned	<p>The Market Survey/Tool Selection provides a quantitative product ranking so that the evaluation team can determine which products to bring into the lab for evaluation.</p> <p>The number of tools selected for evaluation should be based on both the results of the Market Survey/Tool Selection as well as and the resources/timeframe available for the evaluation.</p> <p>Appoint evaluation team member(s) to serve as the point(s) of contact for each product.</p> <p>Before contacting vendors, obtain clear instructions from the project lead and/or sponsor about the type of information that can be communicated to them about the sponsor's identity and mission.</p>

	Request to speak with a technical engineer as well as a sales representative to discuss products' features or capabilities.
Templates and Sample Deliverables (Samples available upon request)	Market Survey/Tool Selection template Sample: Network Intrusion Detection/Prevention Systems (Network IDS/IPS) Market Survey/Tool Selection

4.4 Action: Determine Test Architecture

Description and Activities	Discuss set-up of lab equipment, software, and hardware required for the evaluation. Identify any needs for specialized hardware, software, or other equipment.
Lessons Learned	Create a network diagram illustrating how the lab equipment will be connected.
Templates and Sample Deliverables (Samples available upon request)	Sample: Forensics Evaluation Test Architecture

4.5 Action: Draft High-Level Test Plan

Description and Activities	Prepare a 1-2 page write-up with the following information: <ul style="list-style-type: none"> • Project name, charge number • Points of contact • Project description and objectives • Deliverables • Resources • Test Methodology and Requirements Specify any infrastructure or logistical requirements. For
-----------------------------------	---

	<p>example,</p> <ul style="list-style-type: none"> • Will the evaluation team need to connect to or isolate data from the MITRE Information Infrastructure (MII)? • Will the team perform any product tests in parallel? <p>Complete this write-up using the High-Level Test Plan template.</p> <p>Provide a list of purchase requests to the evaluation team lead if supporting equipment (hardware/software) is required for testing.</p>
Lessons Learned	<p>Write the High-Level Test Plan with as much detail as possible, and as if it is a deliverable for the sponsor or an outside reader. Provide background, introduce the products to be evaluated, and present the testing and scoring methodology. This document will continue to grow and can hopefully become Chapter 2: Test Strategy in the Final Report.</p>
Templates and Sample Deliverables (Samples available upon request)	<p>High-Level Test Plan template</p> <p>Sample: Forensics High-Level Test Plan</p>

4.6 Check Point – Phase 1

Description and Activities	<p>Internal Component: Complete all actions from Phase 1. At a team meeting, discuss the next actions and assign tasks for proceeding with Phase 2.</p> <p>External Component: Engage stakeholders and subject matter experts from the sponsor’s environment with a status brief on the preliminary findings and objectives.</p> <p>Solicit feedback and suggestions on the evaluation approach the products under consideration.</p>
Lessons Learned	<p>Work with the sponsor to select a list of products for full evaluation based on the market survey results, evaluation timeline, and resources available.</p> <p>Below is a suggested briefing format (from Phase 1 Brief</p>

	template): <ul style="list-style-type: none"> • Purpose of Task • Base Assumptions/Key Requirements • Tool Selection • Product Highlights • Product Drawbacks • Background on the evaluation test environment • High-level test architecture/plan • Tool Selection Ranking • Next Steps
Templates and Sample Deliverables (Samples available upon request)	Phase 1 Brief template

5 Phase 2: Test Preparation

After selecting the products to evaluate and obtaining concurrence from the sponsor, the evaluation team works to acquire the evaluation products from the vendors, and any additional infrastructure that is required for testing. This includes signing non-disclosure agreements (NDAs), establishing vendor points of contact, and meeting with the vendor to discuss the test plan. At the same time, the team develops a full set of evaluation criteria that the products will be tested against and any scenario tests that will be performed. The evaluation team then installs the products in the test environment, and engages the vendor as technical questions arise. The team may wish to hold a technical exchange meeting (TEM) to gain further insight and background from subject matter experts.

5.1 Action: Establish Evaluation Criteria, Priorities, and Test Procedures

<p>Description and Activities</p>	<p>Develop technical evaluation criteria and separate the criteria into categories, such as:</p> <ul style="list-style-type: none"> • General functionality • User Interface • Vendor Support • Administration and Management <p>Using the Evaluation Criteria Spreadsheet template, track the following information:</p> <ul style="list-style-type: none"> • Evaluation Criteria Statements • Procedures for How to Test each criterion and the expected outcomes (Test Methodology) • Evaluation Criteria Weights and Category Weights • Score • Comments <p>If the products are tested against any scenarios, provide descriptions of each scenario, its test environment, and how each product will be scored.</p>
<p>Lessons Learned</p>	<p>Number the evaluation criteria so that they can be referenced easily throughout the evaluation and in the final report.</p> <p>Ensure that all aspects and objectives of the technology evaluation are addressed. Clearly state <i>what</i> is being tested, <i>how</i></p>

	<p>it is being tested, and the expected outcome. The wording of the evaluation criteria must be precise so that:</p> <ul style="list-style-type: none"> • Each member of the evaluation team understands and has the same interpretation of the criteria and test procedures. • If an evaluation criterion is vague, an outsider (vendor, stakeholder) may misinterpret it in the final report and challenge the associated product scores. <p>It is critical that the evaluation team be able to defend their tests and results with documented statements and procedures from Phase 2.</p> <p>Establish weights for the individual evaluation criteria and the evaluation categories. For a thorough explanation of established/approved scoring techniques, see Section 3. The Reference Comparison method is recommended for eliciting evaluation criteria weights. The Paired Comparison method is recommended for eliciting evaluation category weights.</p> <p>Consider dividing the test results into two test phases: Evaluation Criteria Testing (Phase 1) and Scenario Testing (Phase 2) to distinguish between evaluation criteria, which are usually single steps, and scenarios, which cover a number of steps.</p>
<p>Templates and Sample Deliverables (Samples available upon request)</p>	<p>Evaluation Criteria Spreadsheet template</p> <p>Sample: Forensics Evaluation Criteria</p> <p>Sample: Forensics Evaluation Scenario and Scenario Scoresheet</p>

5.2 Action: Perform Government Requirements’ Mapping

<p>Description and Activities</p>	<p>Identify key government policy documents or guidance (e.g., DCID 6/3 for Information Systems within the Intelligence Community) that may necessitate additional evaluation criteria.</p> <p>Cross-reference the Evaluation Criteria against these documents to verify that the criteria cover all government requirements and to document the mapping.</p>
--	--

Lessons Learned	Completing this mapping may not be straightforward, as certain criteria may not map directly to government requirements. Therefore, assign more than one team member to perform this action to ensure a consensus has been reached in the final results.
Templates and Sample Deliverables (Samples available upon request)	None

5.3 Action: Enhance and Finalize Test Plan

Description and Activities	<p>Using the High-Level Test Plan and Project Summary, write a more detailed, 2-3 page Test Plan to include:</p> <ul style="list-style-type: none"> • Short descriptions of the tools selected for evaluation (including version numbers and all components) • Any data sources (sample files, network data, etc.) to be used in testing • A network diagram of the test architecture • Additional hardware and software to be used for testing • Infrastructure requirements (e.g., access to the MII or a separate, closed network) • Categories covered by the evaluation criteria • Description of any scenarios to be tested and how testing will be performed • The scoring methodology that will be used (see Section 2.2.2)
Lessons Learned	<p>Write the Test Plan as if it is a deliverable for the sponsor or an outside reader. Provide background, introduce the products to be evaluated, and present the testing and scoring methodology. With minor modifications, it should be possible to insert the Test Plan directly as Chapter 2: Test Strategy in the Final Report. This will alleviate a lot of pain later!</p>

Templates and Sample Deliverables (Samples available upon request)	High-Level Test Plan template Sample: Forensics Finalized Test Plan
--	--

5.4 Action: Acquire Necessary Hardware and Software

Description and Activities	<p>For each product selected for evaluation, contact the product vendors and discuss the following:</p> <ul style="list-style-type: none"> • Explain that their product has been selected to be part of an evaluation at MIRE, for the government • Review the purpose and scope of the evaluation (list some of the evaluation categories, scenarios that may be tested, clarify any details that are <i>outside</i> the scope of the evaluation) • Verify contact information for the sales representative and technical engineer points of contact who will work with the evaluators during the evaluation • Explain that prior to delivering a final report to the sponsor, the evaluation findings and results will be reviewed with the vendor • Obtain any non-disclosure agreements (NDAs) and provide these to MITRE Contracts for signature. MITRE technical staff members are not authorized to sign NDAs.
Lessons Learned	<p>As stated in Action 4: Market Survey/Tool Selection, obtain clear instructions from the project lead and/or sponsor about the type of information that can be communicated to vendors verbally and electronically about the sponsor’s identity and mission.</p> <p>Evaluations provide valuable feedback to vendors about their products and may ultimately result in a government purchase for a significant customer base. As a result, vendors are asked to provide evaluation products free of charge. MITRE/sponsors typically do not pay for this equipment.</p>

	<p>Request all equipment/license keys for the duration of the evaluation, so that the evaluation team can tests repeat or verify tests until the final report is delivered.</p> <p>If time affords, resist vendor offers to set up the evaluation equipment for testing. Installation and configuration should be included as part of the evaluation, therefore, it is important that the evaluation team set up the equipment in the lab on their own. Products should also be configured in accordance with their sponsor’s environment. During a short-term evaluation, however, it may be better for the vendor to set up the equipment quickly.</p>
<p>Templates and Sample Deliverables (Samples available upon request)</p>	None

5.5 Action: Hold Technical Exchange Meeting (TEM) (optional)

<p>Description and Activities</p>	<p>It may be useful to have a Technical Exchange Meeting (TEM) with subject matter experts, stakeholders, and the sponsor to share information about the problem or technology space. A TEM may consist of a series of presentations and discussions that can influence how the problem is viewed, the weights assigned to the evaluation criteria, and to provide guidance for MITRE and the sponsor on their future involvement in the subject area.</p>
<p>Lessons Learned</p>	<p>A TEM provides an opportunity for informal communication among project stakeholders, the evaluation team, and the sponsor during the project.</p> <p>Key success factors for a TEM include</p> <ul style="list-style-type: none"> • Identifying and inviting the right people • Inviting the sponsor • A solicitation for presentations that are relevant • Participation among subject matter experts who may not be familiar with the project and can offer a new, alternate

	perspective
Templates and Sample Deliverables (Samples available upon request)	None

5.6 Check Point – Phase 2

Description and Activities	<p>Internal Component: Incorporate feedback received from the sponsor and stakeholders on the evaluation criteria and weights.</p> <p>Complete all evaluation equipment delivery, set-up, and configuration. Prepare a status briefing for the sponsor and stakeholders that covers the Test Plan and Evaluation Criteria.</p> <p>External Component: Provide a status brief for the sponsor, and stakeholders/subject matter experts from the sponsor’s environment. Review the evaluation criteria categories, the criteria themselves, and the assigned weights. Request input and feedback stakeholders on any criteria and/or weight adjustments that should be considered.</p> <p>Below is a suggested briefing format (from Phase 2 Brief template):</p> <ul style="list-style-type: none"> • Agenda • Purpose • Current Status – Equipment Acquisition and Set-Up • Overview of Finalized Evaluation Criteria • Overview of Finalized Evaluation Weights • Overview of Test Plan • Timeline and Next Steps
Lessons Learned	<p>Provide soft copy and hard copy versions of the evaluation criteria to the sponsor and stakeholders involved.</p> <p>Send a follow-up email to stakeholders requesting input on the</p>

	evaluation criteria 1-2 weeks before they are finalized.
Templates and Sample Deliverables (Samples available upon request)	Phase 2 Brief template

6 Phase 3: Testing, Results, and Final Report

In this phase, the evaluation team tests and scores the products against all of the test criteria. The team must ensure that testing for each product is performed under identical conditions, and must complete a full crosswalk of the scores for each product requirement after testing to ensure scoring consistency. Following the crosswalk, evaluation team members conduct individual meetings with each vendor to review their findings, correct any misunderstandings about their product's functionality, and retest if necessary. The team produces a final report that incorporates the evaluation results and any supporting information.

6.1 Action: Conduct Testing and Compile Results

Description and Activities	<p>Perform all functional testing and proceed to any scenario testing. Score the product against every criterion and fully document all comments in the comment field <i>during</i> testing so that no information is lost or forgotten. The evaluation criteria results will be included in the final report.</p> <p>The evaluation team must ensure that each product tested receives the same data and is tested in the same environment. In addition, data collection must be performed uniformly for all products. NOTE: Test results will be repeatable only to the extent that the modeled scenario and data can be reproduced.</p>
Lessons Learned	<p>To maintain objectivity, consider testing each product according to the 2-person rule—evaluate each product in pairs, and iterate through different pair combinations across the products. (i.e., two people who evaluate one product work with new partners to evaluate subsequent products).</p> <p>Score products according to one of the established scoring methods presented in Section 2.2.2.</p> <p>Contact the vendor if problems are encountered during testing. Discuss the issues, request troubleshooting support, and correct any misunderstandings.</p> <p>As a team, agree on a policy for applying any patches or updates that are issued after products are installed in the lab. Teams may wish to apply updates that help to improve the product if testing can still be completed on time</p>

Templates and Sample Deliverables (Samples available upon request)	None
--	------

6.2 Action: Perform Crosswalk

Description and Activities	<p>Meet as a team, and review the evaluation criteria in order and the assigned scores for each product. The purpose of the crosswalk is to ensure that the products are scored consistently.</p> <p>Clarify any scores of “1” (criteria that are partially met) with an explanation in the Evaluation Criteria comment field. These comments may or may not be included in the final report, but the team must document somewhere their rationale for awarding partial credit.</p>
Lessons Learned	<p>Set aside several hours over multiple days to perform the crosswalk. While it is a lengthy process, the crosswalk is critical to ensure that the team discusses and normalizes the product scores for each criterion.</p>
Templates and Sample Deliverables (Samples available upon request)	None

6.3 Action: Share Results with Vendors

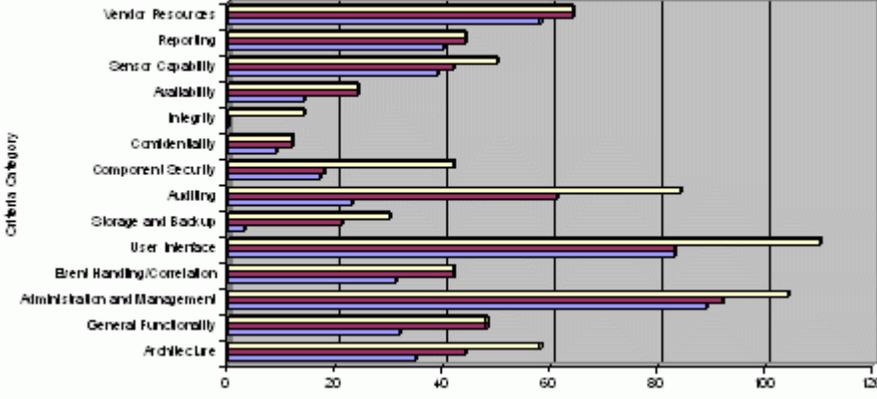
Description and Activities	<p>Following the crosswalk, schedule individual vendor meetings to review the evaluation findings.</p> <p>For each product, discuss the overall strengths and weaknesses and obtain concurrence from the vendor that the team’s findings accurately capture the product’s capabilities.</p> <p>Provide vendors with an opportunity to correct any misunderstandings about their product’s functionality, and retest if necessary.</p>
-----------------------------------	---

	Discuss potential product improvements and/or changes from the perspective of the sponsor's environment.
Lessons Learned	<p>The sponsor is the owner of the evaluation itself, and as a result, MITRE is obligated to protect specific objectives, requirements, and intentions throughout the evaluation. In addition, NDAs that were signed with each of the vendors prevents the evaluation team from sharing results with their competitors.</p> <p>For these reasons, when reviewing evaluation results with vendors, do not:</p> <ul style="list-style-type: none"> • Discuss any other product's performance in the evaluation. • Reveal the weights of individual evaluation criteria and/or category weights • Provide vendors with copies of their product's results (unless otherwise directed by the sponsor) <p>During the vendor briefing, the evaluation team should:</p> <ul style="list-style-type: none"> • Ensure both members of the 2-person evaluation team that evaluated the product are present • Review major strengths and weaknesses found in the product • Discuss overall impressions • Discuss any lingering problems encountered during testing • Allow the vendor to correct any misunderstandings
Templates and Sample Deliverables (Samples available upon request)	None

6.4 Action: Deliver Final Report

Description and	Structure the final report as follows for both technology and product evaluations:
------------------------	--

Activities	<ul style="list-style-type: none"> • Executive Summary • Table of Contents • Introduction <ul style="list-style-type: none"> ○ Background ○ Purpose ○ Organization of Document • Test Preparation • Findings -- the following sections for each product: <ul style="list-style-type: none"> ○ Strengths ○ Weaknesses • Recommendations and Conclusions • References • Appendices <ul style="list-style-type: none"> ○ Test Results ○ Evaluation Criteria ○ Glossary and Acronyms ○ Test Data (if applicable)
Lessons Learned	<p>Create charts and graphs to capture the overall evaluation results. Capture scores and performance for each product so the reader can visualize the results of the evaluation. Below is a sample chart from an Intrusion Detection Evaluation:</p>

	 <p>Dedicate a section to each product evaluated. Call out its strengths and weaknesses and ensure that these are the same strengths and weaknesses reflected in the numerical evaluation results. That is, identify the categories in which the product scored highest (or stood out above other products). Identifying a product’s strengths/weaknesses based on the team’s recollection of the evaluation is not a reliable tool.</p>
<p>Templates and Sample Deliverables (Samples available upon request)</p>	<p>Samples available upon request</p>

6.5 Check Point – Phase 3

<p>Description and Activities</p>	<p>Internal Components: Complete all testing, including the initial evaluation and scoring, as well as any further testing required as a result of the crosswalk and/or vendor meetings.</p> <p>Complete all vendor meetings to discuss the evaluation findings and overall results for the final report.</p> <p>Draft final report, complete MITRE internal peer review, and deliver report to the sponsor.</p>
--	---

	External Component (optional): Prepare final status brief for sponsor and key stakeholders to review the evaluation objectives and goals, and present preliminary findings/recommendations.
Lessons Learned	Ensure that all vendor concerns are discussed and an agreement is reached before any controversial statements are written in the final report.
Templates and Sample Deliverables (Samples available upon request)	None

7 Acknowledgments

This work was completed with funding from the MITRE Systems Engineering Process Office (SEPO), the Office of the Director of National Intelligence (ODNI) Chief Information Officer (CIO), and the G020 - Information Security division. The following people provided much guidance and assistance during this project: Chris Do, Dale Johnson, Robin Medlock, Michael O'Connor, Bill Neugent, Greg Stephens, Jake Ulvila, John Vasak, Lora Voas, and Brian White.

8 References

URLs are valid as of the date of this document.

- [1] J. Butler, D. J. Morrice, and P.W. Mullarkey. 'A Multiple Attribute Utility Theory Approach to Ranking and Selection', *Management Science*, 47/6:800-816. (2001).
- [2] T. Edmunds. 'Multiattribute Utility Analysis (MAU) to Support Decisions', (presentation), Systems and Decision Sciences Technology, Lawrence Livermore National Laboratory. (2001).
- [3] W. Edwards. 'SMART and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement', *Organizational Behavior and Human Decision Processes*, 60:306-325. (1994)
- [4] E. H. Forman and S. I. Gass. 'The Analytic Hierarchy Process—An Exposition', *Operations Research, INFORMS*, 49/4:469-486. (2001).
- [5] R. Haas and O. Meixner. 'An Illustrated Guide to the Analytic Hierarchy Process', Institute of Marketing and Innovation, University of Natural Resources and Applied Life Sciences, Vienna. <<http://www.boku.ac.at/mi/ahp/ahptutorial.pdf>> (2007).
- [6] R. L. Keeney. 'Common Mistakes in Making Value Trade-Offs', *Operations Research, INFORMS*, 50/6:935-945. (2002).
- [7] Z. F. Lansdowne and B. W. Lamar. 'An On-Line Survey of Portfolio Selection Methodologies', Center for Enterprise Modernization, The MITRE Corporation. (2003).
- [8] B. Miller, K. See, and N. Tronick. 'MITRE Center for Enterprise Modernization Project Leader Handbook.' Draft, version 2.0. The MITRE Corporation. (2006).
- [9] 'Modelling and Decision Support Tools', Institute for Manufacturing, University of Cambridge. <<http://www.ifm.eng.cam.ac.uk/dstools/#3>>. (2007).
- [10] T. L. Saaty. "Priority Setting in Complex Problems", *IEEE Transactions on Engineering Management*, 30/3140-155. (1983).

- [11] T. L. Saaty. 'The analytic hierarchy process: Some observations on the paper by Apostolou and Hassell', Journal of Accounting Literature. (1994).
<http://www.findarticles.com/p/articles/mi_qa3706/is_199401/ai_n8722119>
- [12] 'A Systems Approach to Project Management', Cambridge Consulting. MITRE Institute Course Project Management Boot Camp. (2007)
- [13] 'Test & Evaluation Handbook for C2 Systems', Draft, ESC Test and Evaluation Directorate, The MITRE Corporation. (1998).
- [14] J. W. Ulvila, et al. 'A Framework for Information Assurance Attributes and Metrics', Technical Report 01-1. Decision Sciences Associates, Inc. (2001)

Appendix A Acronym and Definition List

Technology evaluation	An evaluation of multiple products from the same technology area
Product evaluation	An evaluation of a single product for use in a sponsor's environment
Evaluation scenario	Procedures designed to test a product's performance in a particular application or situation
Evaluation criteria	Functional requirements and features that products are tested against in an evaluation
TEM	Technical Exchange Meeting
IC TIC	Intelligence Community Test and Integration Center
STEP	Standardized Technical Evaluation Process

Appendix B STEP Templates

Section	Action	Template
§ 3.1	Conduct Preliminary Scoping	Preliminary Scoping
§ 3.2	Scoping with Government Sponsor	Scoping with Sponsor
§ 3.3	Perform Market Survey/Tool Selection	Market Survey/Tool Selection
§ 3.4	Determine Test Architecture	
§ 3.5	Draft High-Level Test Plan	High-Level Test Plan
§ 3.6	Check Point – Phase 1	Phase 1 Brief
§ 4.1	Establish Evaluation Criteria, Priorities & Test Procedures	Evaluation Criteria Spreadsheet
§ 4.2	Perform Government Requirements' Mapping	
§ 4.3	Enhance and Finalize Test Plan	High-Level Test Plan
§ 4.4	Acquire Necessary Hardware and Software	
§ 4.5	Hold Technical Exchange Meeting (TEM) (optional)	
§ 4.6	Check Point – Phase 2	Phase 2 Brief
§ 5.1	Conduct Testing and Compile Results	
§ 5.2	Perform Crosswalk	
§ 5.3	Share Results with Vendors	
§ 5.4	Deliver Final Report	
§ 5.5	Check Point – Phase 3	
§ 6.1	Purchase Selected Product	
§ 6.2	Support Sponsor with Integration and Deployment	

Insert classification, (e.g., UNCLASSIFIED//FOUO)

B.1 Preliminary Scoping Template

Created by: _____

Date: _____

Product Name	<product name>
Vendor	
URL	
Company Headquarters Location	
Company Size	
Foreign Owners or Investors?	
Pricing	
Contacts	
Commercial/Open Source	
Does it Integrate with Other Products?	
Current Users (Any government agencies?)	
Operating Platform/OS	
Target OS/File Structure	
<specific question>	
<specific question>	
<specific question>	
Other Key Features	

Insert classification (e.g., UNCLASSIFIED//FOUO)

B.2 Scoping with Sponsor Template

1. <Question>

<Explanation>

Assumption to be confirmed with sponsor: <Assumption>

Sample Question:

1. Are there any preferences on the type of platform used?

Windows, Linux, Solaris, BSD, or some customized system. What about versions of each platform?

Assumption to be confirmed with sponsor: No preference on platform.

Other Possible Scoping Questions:

- 1. What vendors/products should we absolutely include in this evaluation?**
- 2. Is there a specific capability that is required?**
- 3. What is the timeframe for the evaluation?**
- 4. What is the background/motivation for this evaluation?**
- 5. Does the system need to integrate with any external systems?**
- 6. Is there an upper price point that we should consider?**
- 7. Are we looking at a single solution or should we consider products that complement one another?**
- 8. Who is a good resource to talk to as we progress with the evaluation?**
- 9. Are there government requirements that must be met for certification and accreditation?**
- 10. Which organizations should be involved throughout this evaluation?**

Insert classification (e.g., UNCLASSIFIED//FOUO)

B.3 Market Survey/Tool Selection Template

#	Weighted Factors	Weight	<product>	<product>	<product>	<product>	<product>
	<Category 1>	0					
1		0	0	0	0	0	0
2		0	0	0	0	0	0
3		0	0	0	0	0	0
4		0	0	0	0	0	0
	<Category 2>	0					
5		0	0	0	0	0	0
6		0	0	0	0	0	0
7		0	0	0	0	0	0
8		0	0	0	0	0	0
9		0	0	0	0	0	0
10		0	0	0	0	0	0
11		0	0	0	0	0	0
12		0	0	0	0	0	0
	General						
13		0	0	0	0	0	0
14		0	0	0	0	0	0
15		0	0	0	0	0	0
16		0	0	0	0	0	0
	Total Score		0	0	0	0	0

Insert classification (e.g., UNCLASSIFIED//FOUO)

B.4 High-Level Test Plan Template

Project Name name

Project Number number

Contacts

- Sponsor:
- Project Lead:
- Researchers:

Project Description/Initial Objectives

- text

Test Strategy/Methodology

- text

Deliverables

- list

Resources

- Hardware
- Software
- Books
- Training/Committees/TEMs
- Newsgroups
- Project Documents

B.5 Phase 1 Brief Template



1



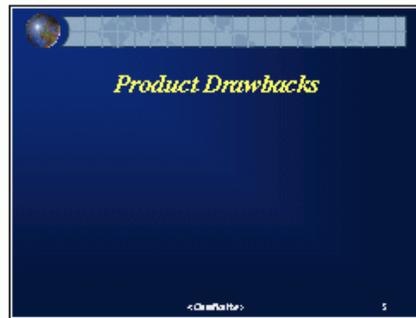
2



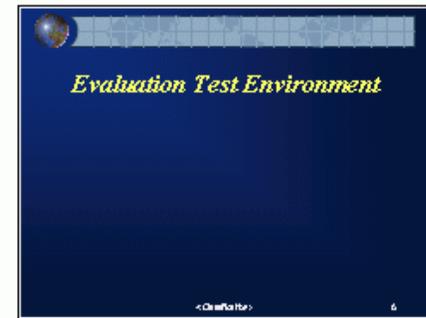
3



4



5



6



7



8



9

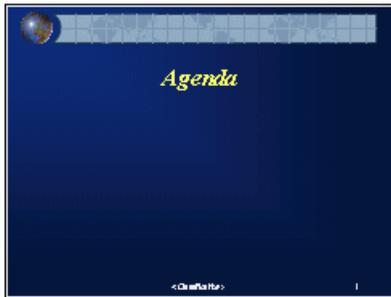
Insert classification (e.g., UNCLASSIFIED//FOUO)

B.6 Evaluation Criteria Spreadsheet

#	Evaluation Criteria	Test Description and How to Test	Weight	<product name>	Comments				
1.0 Category 1									
1.1			0	0	0	0	0	0	
1.2			0	0	0	0	0	0	
1.3			0	0	0	0	0	0	
1.4			0	0	0	0	0	0	
2.0 Category 2									
2.1			0	0	0	0	0	0	
2.2			0	0	0	0	0	0	
2.3			0	0	0	0	0	0	
2.4			0	0	0	0	0	0	
2.5			0	0	0	0	0	0	
2.6			0	0	0	0	0	0	
2.7			0	0	0	0	0	0	
2.8			0	0	0	0	0	0	
3.0 Category 3									
3.1			0	0	0	0	0	0	
3.2			0	0	0	0	0	0	
3.3			0	0	0	0	0	0	
3.4			0	0	0	0	0	0	
Total Score									
				0	0	0	0	0	

Insert classification (e.g., UNCLASSIFIED//FOUO)

B.7 Phase 2 Brief Template



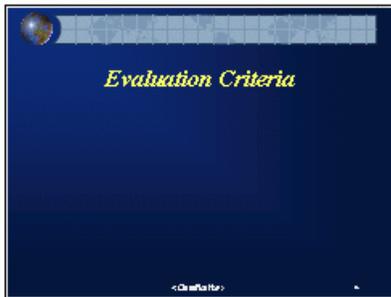
1



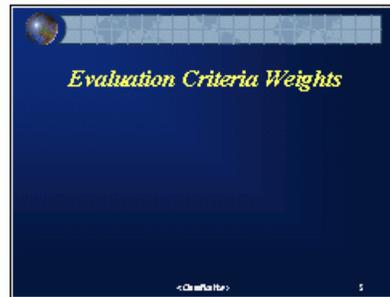
2



3



4



5



6



7