Guide for Understanding and Implementing Defense Experimentation (GUIDEx)

Experimentation offers a unique means to support the development and transformation of allied forces by advancing our knowledge of the complex networked systems and capabilities likely to be fielded in the near future.

The process of organized and integrated experimentation accelerates the development of improved tools, tailored skills, new processes, and alternative technologies.

The growing importance of experimentation motivated TTCP's Joint Systems and Analysis Group (JSA) to establish Action Group 12 on Methods and Approaches for Warfighting Experimentation in 2002. The work of AG-12 over three years culminated in a 350-page guide for defense experimentation. GUIDEx structures the best practices for experimentation around 14 Principles to ensure that allied defense experimentation programs are genuinely able to support the evolution of the force capabilities of the future. For the benefit of readers a set of real-world Case Studies is provided to illustrate the 14 Principles in practice. They also provide further material for devising a way ahead for accelerating the acquisition of knowledge to maintain a leading advantage in military capabilities.

Although this guide has been written mainly for the practitioners and designers of defense experimentation, we hope that it will stimulate better communication among military officers, government officials and the defense scientific communities of the allied nations on all matters associated with defense experimentation. Additionally, the experimentation Principles described in this guide apply to other large enterprises and multiple agency operations, for example in homeland security.

GUIDEx can be downloaded from the TTCP website.

Who should read GUIDEx?

- Those who ask force capability questions and act on the answers.
- Those who decide how the force capability question is to be addressed and what methods are to be used.
- Those who design, execute, and interpret defense (warfighting) experiments.
- Those engaged in Operational Test & Evaluation (OT&E).
- All those for whom experimentation matters!

Aim and purpose of TTCP

The aim of TTCP is to foster cooperation within the science and technology areas needed for conventional (*i.e.*, non-atomic) national defense. Its purpose is to enhance national defense and reduce costs. More information on TTCP can be found on its public website at **http://www.dtic.mil/ttcp/**

The thesis of GUIDEx is that robust experimentation methods from the sciences can be adapted and applied to military experimentation and will provide the basis for advancements in military effectiveness in the transformation process.

Funds for printing GUIDEx were provided by the Canadian Forces Experimentation Centre, Ottawa, Canada

GUIDEx includes a discussion of the logic for experimentation.

The logic of experimentation can be expressed by a mnemonic in the numbers 2,3,4, 5 and 21.

The number "2" represents the **two components/sides** of an experiment hypothesis – the left-hand side and the right-hand side, the "if" side and the "then" side. **If** new solution **A** is used, **then** effect **B** indicates that the operational problem might be solved.

The number "3" shows that there are **three logical steps** to resolve hypotheses:

1. The first logical question is whether the proposed solution **A**, the left-hand side of the hypothesis, was adequately represented in the experiment.
2. The second question is whether the experimenter was able to observe that the experiment produced evidences (effect **B**) in an objective manner that the problem to be solved was, in fact, solved.
3. Given that the proposed solution was adequately represented and given that progress was observed in solving the problem, the third logical question concerns whether the observed problem resolution was due to the proposed solution. Does **A** really cause **B**?

The number "4" indicates that good (valid) Defense experiments should be designed to meet the **four validity requirements**:

1. Ability to employ the new capability
2. Ability to detect change
3. Ability to isolate the reason for change
4. Ability to relate results to actual operations

The number "5" confirms that all experiments – large or small, field or laboratory, military or academic, applied or pure – consist of **five components**:

1. The treatment, the possible cause **A**, is a capability or condition that may influence warfighting effectiveness.
2. The possible effect **B** of the treatment is the result of the trial, an increase or decrease in some aspect of warfighting effectiveness.
3. The experimental unit executes the possible cause and produces an effect.
4. The trial is one observation of the experimental unit under treatment **A** or under the alternative **~A** to the new capability to see if effect **B** occurred or not, and includes all of the contextual conditions under which the experiment is executed.
5. The analysis phase of the experiment compares the results from one trial to a different trial.

Cook and Campbell's threats to validity can be distilled down to 21 threats to defense experiments. These threats can be arrayed within a two-dimensional matrix to better understand the actions the experimenter can take to counter these threats. In the illustration, the 21 threats to validity are arrayed with respect to the four experiment validity requirements and the five experiment components.

All good experiment practices are counters, or antidotes, to the 21 threats to experiment validity. A good experiment plan should show how each of the 21 threats has been accounted for and countered.

GUIDEx provides good experiment design practices to counter each of the 21 threats.

21 Threats to a Good Defense Experiment

4 Experiment Requirements

5 Experiment Components	1 Ability to **Use** Capability	2 Ability to **Detect** Change	3 Ability to **Isolate** Reason for Change: Single Group	3 Ability to **Isolate** Reason for Change: Multiple Groups	4 Ability to **Relate** Results to Operations
1 Treatment	1. Capability not workable: Do the hardware and software work?	5. Capability variability: Are systems (hardware and software) in use in like trials the same?	11. Capability changes over time: Are there system (hardware or software) or process changes during the test?	N/A	18. Nonrepresentative capability: Is the experimental surrogate functionally representative?
2 Players	2. Player non-use: Do players have the training and tactics, techniques and procedures (TTPs) to use the capability?	6. Player variability: Do individual operators/units in like trials have similar characteristics?	12. Player changes over time: Will the player unit change over time?	15. Player differences: Are there differences between groups unrelated to the treatment?	19. Nonrepresentative players: Is the player unit similar to the intended operational unit?
3 Effect	3. No potential effect in output: Is the output sensitive to capability use?	7. Data collection variability: Is there a large error variability in the data collection process?	13. Data collection changes over time: Are there changes in instrumentation or manual data collection during the experiment?	16. Data collection differences: Are there potential data collection differences between treatment groups?	20. Nonrepresentative measures: Do the performance measures reflect the desired operational outcome?
4 Trial	4. Capability not exercised: Do the scenario and Master Scenario Event List (MSEL) call for capability use?	8. Trial conditions variability: Are there uncontrolled or unmonitored changes in trial conditions for like trials? Look for intervening variables not recorded.	14. Trial conditions change over time: Are there changes in the trial conditions (such as weather, light, start conditions, and threat) during the experiment?	17. Trial conditions differences: Are the trial conditions similar for each treatment group?	21. Nonrepresentative scenario: Are the Blue, Green, and Red conditions realistic?
5 Analysis	N/A	9. Low statistical power: Is the analysis efficient and the sample sufficient? 10. Violation of statistical assumptions: Are the correct analysis techniques used and error rate reduced?	• The purpose of an experiment is to verify that **A** causes **B**. • A valid experiment allows the conclusion, **A** causes **B**, to be based on evidence and sound reasoning... — by reducing or eliminating the 21 known threats to validity.		

GUIDEx is organized along the following 14 Principles for effective experimentation. They are grouped under three dominant topics or themes as it follows:

Designing Valid Experiments

1. Defense experiments are uniquely suited to investigate the cause-and-effect relationships underlying capability development.
2. Designing effective experiments requires an understanding of the logic of experimentation.
3. Defense experiments should be designed to meet the four validity requirements.

Integrated Analysis and Experimentation Campaigns

4. Defense experiments should be integrated into a coherent campaign of activities to maximize their utility.
5. An iterative process of problem formulation, analysis and experimentation is critical to accumulate knowledge and validity within a campaign.
6. Campaigns should be designed to integrate all three scientific methods of knowledge generation (studies, observations and experiments).
7. Multiple methods are necessary within a campaign in order to accumulate validity across the four requirements.

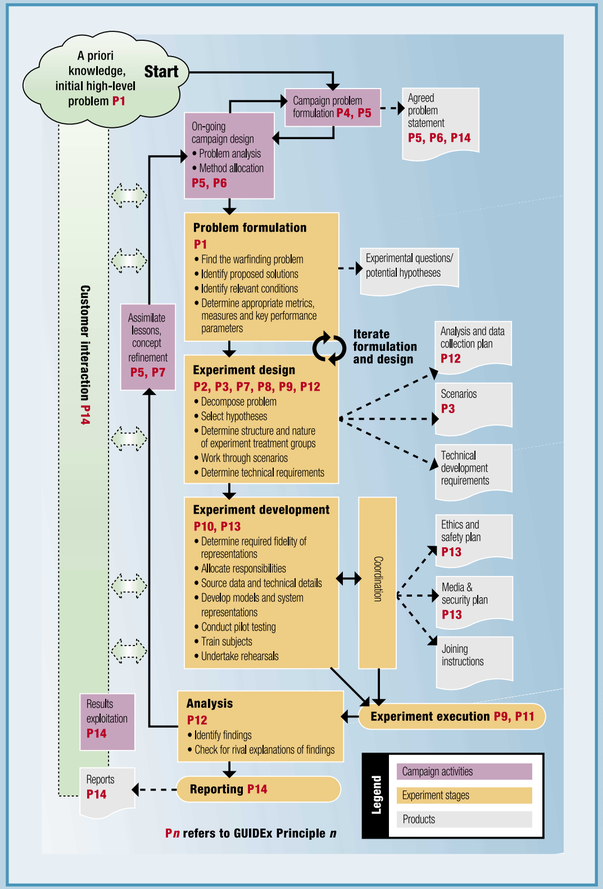
Considerations for Successful Experimentation

8. Human variability in defense experimentation requires additional experiment design considerations.
9. Defense experiments conducted during collective training and operational test and evaluation require additional experiment design considerations.
10. Appropriate exploitation of modeling and simulation is critical to successful experimentation.
11. An effective experimentation control regime is essential to successful experimentation.
12. A successful experiment depends upon a comprehensive data analysis and collection plan.
13. Defense experiment design must consider relevant ethical, environmental, political, multinational, and security issues.
14. Frequent communication with stakeholders is critical to successful experimentation.

In order to help practitioners in applying the GUIDEx principles to address their specific problems, the following flowchart was developed. This is by no means a prescriptive recipe for perfect experimentation, but an attempt to lay out the chronological sequence for experiment and campaign related activities and to show the iterations and linkages between various stages of the experimentation process. GUIDEx encourages that the specific application of Principles to a given problem should be tailored according to the scale and nature of the issue under investigation.

The color code of the flowchart separates the integrated analysis and experimentation campaign activities (in purple) from the specific individual experiment stages (in orange). The grey indicates the products of the experimentation process, while green shows the customer or stakeholder interactions. The flowchart starts from initial discussions with the customer(s) through an integrated analysis and experimentation approach that ensures co-evolution of a campaign plan for improved execution and exploitation of each experiment, study and result in order to better impact sought future force capabilities.

Experiment and Campaign Planning Flowchart



Art Direction by ADM (PA) Creative Services CS05-0345

Guide for Understanding and Implementing Defense Experimentation

GUIDEx

The Technical Cooperation Program