INSTITUTE FOR DEFENSE ANALYSES

DATAWorks 2022: A New Method for Planning Full-Up System-Level (FUSL) Live Fire Tests

Mark A. Couch, Project Leader

Lindsey D. Butler
Matthew R. Avery
William G. Gardner

March 2022

Public release approved. Distribution is unlimited.

IDA Document NS D-32966

Log: H 2022-00043

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, C9107, “Live Fire Methodology Advancements,” for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of David K. Grimm and Mark A. Couch from the Operational Evaluation Division.

For more information:
Mark A. Couch, Project Leader
mcouch@ida.org • (703) 845-2530

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-32966

DATAWorks 2022: A New Method for Planning Full-Up System-Level (FUSL) Live Fire Tests

Mark A. Couch, Project Leader

Lindsey D. Butler
Matthew R. Avery
William G. Gardner

Executive Summary

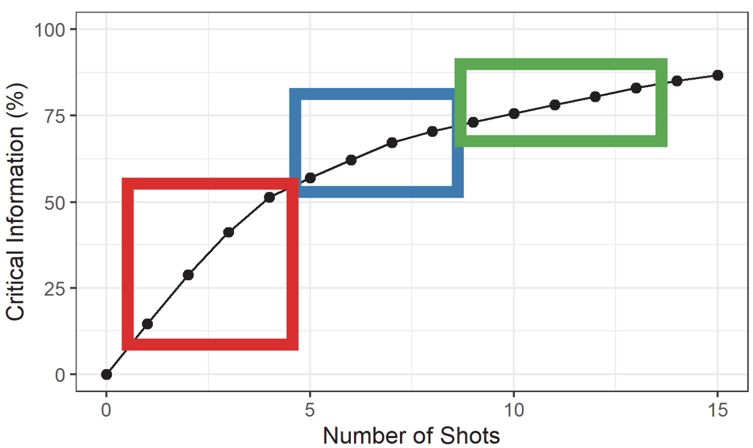
Planning Full-Up System-Level (FUSL) Live Fire tests is a complex process that has historically relied heavily on subject matter expertise. In particular, there is no established method or process to determine the appropriate number of FUSL tests necessary for a given program. IDA funded a project through its Central Research Proposal to develop a novel method that is analogous to the Design of Experiments process that is used to determine the scope of Operational Test events.

Currently, FUSL test planning relies on subject matter experts (SMEs) to identify both the total number of FUSL events and the specific conditions for each event. SMEs identify the most critical shots to test by considering the potential for troop casualties, degradation of mission essential functions, and the operational cost to recover and repair the vehicle. To reduce the overall number of tests, SMEs also ensure all FUSL tests are dissimilar enough so that we learn new information with each FUSL event that we conduct. Our proposed methodology applies the same principles as the current FUSL test planning process, but collects inputs from the SMEs in a more structured manner.

First, SMEs define all potential FUSL shots. For each potential shot, SMEs estimate the severity of that shot, the uncertainty of that severity estimate, and the similarity of that shot to all other potential shots. Second, we developed a numerical optimization algorithm that uses the SME inputs to generate a prioritized list of FUSL events and a corresponding plot of the total information gained with each successive shot.

Together, these outputs can help analysts determine the adequate number of FUSL tests for a given program and ensure that the most critical shots are prioritized. We illustrate this process with an example on a notional ground vehicle. Future work, including sensitivity studies, algorithm improvements, and retrospective studies on previous programs of record, is necessary prior to implementation on a new program of record.

A New Method for Planning Full-Up System-Level (FUSL) Live Fire Tests



Lindsey Butler

Matt Avery

Will Gardner

April 2022

Number	Variant	Threat Severity	Shot Line
1	Variant A	Beyond Threshold	Driver
2	Variant B	Beyond Threshold	Engine
3	Variant C	Beyond Threshold	Commander
4	Variant D	Beyond Threshold	Undertrack
5	Variant E	Threshold	Commander
6	Variant A	Threshold	Driver
7	Variant B	Threshold	Engine
8	Variant E	Beyond Threshold	Undertrack
9	Variant D	Beyond Threshold	D Crew
10	Variant C	Beyond Threshold	C Crew
11	Variant A	Beyond Threshold	A Crew
12	Variant B	Beyond Threshold	B Crew
13	Variant E	Beyond Threshold	E Crew
14	Variant D	Beyond Threshold	Commander
15	Variant A	Beyond Threshold	Engine

Institute for Defense Analyses

730 East Glebe Road ● Alexandria, Virginia 22305

What is Live Fire Testing?

BLUF

- We will present an approach for sizing Full-Up System-Level (FUSL) Live Fire testing, that aims to help analysts answer the question:
  - How many FUSL events are necessary for a program?
- We developed a quantitative optimization algorithm for determining FUSL test scope that formalizes the FUSL test planning process
  - This approach is analogous to the type of DOE used in OT&E
- We will walk through a notional example to illustrate proof of concept
- We identify future work that is required before this approach could be employed in practice

Acronyms: BLUF – Bottom Line Up Front; DOE – Design of Experiments; LFT&E – Live Fire Test and Evaluation; OT&E – Operational Test and Evaluation

Design of Experiments is both a general approach to experimentation and a particular set of algorithms and tools

Identify test objectives and metrics

Define factor space

Posit relationships between factors (interactions, etc.)

Choose a set of test points that fulfills some criteria:

- Orthogonality
- D-Optimality
- Model coverage
- Factor power

While typical DOE tools (power analyses) may not be appropriate for FUSL test design, the general approach is to:

Identify test objectives

Define the space of potential shots

Posit relationships between “factors” and outcomes

Choose test points to optimize for information gained

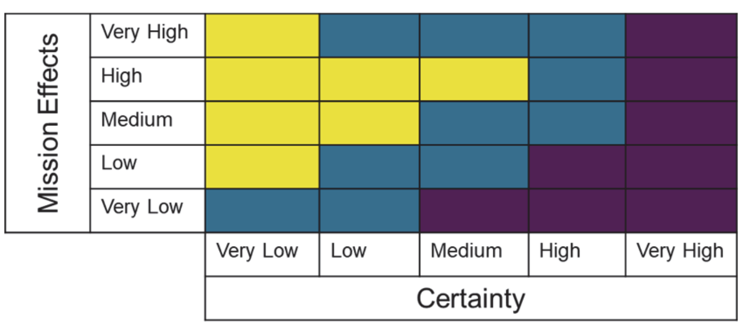
Acronyms: DOE – Design of Experiments; FUSL – Full-Up System-Level

Overview of approach for early FUSL test planning

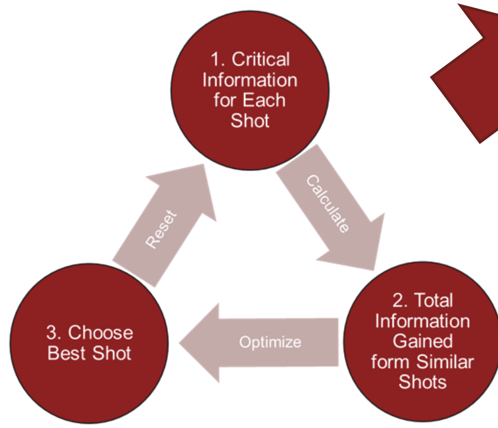
Identify All Potential Shots

Factor	Levels
Variant	Variant A, Variant B, Variant C, Variant D, Variant E
Threat Category	Underbody Blast (UBB)
Threat Severity	Threshold-level, Beyond Threshold-level, Objective-level
Location	Driver, Commander, Crew Compartment, Engine, Undertrack

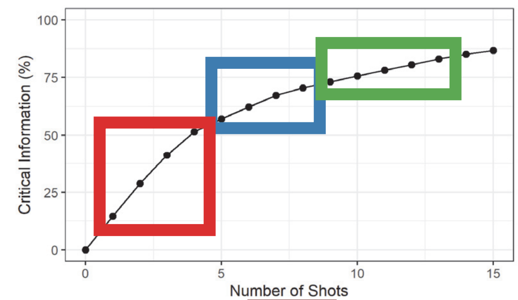
Assign Priority



Algorithmically Find Shots that Provide the Most Information



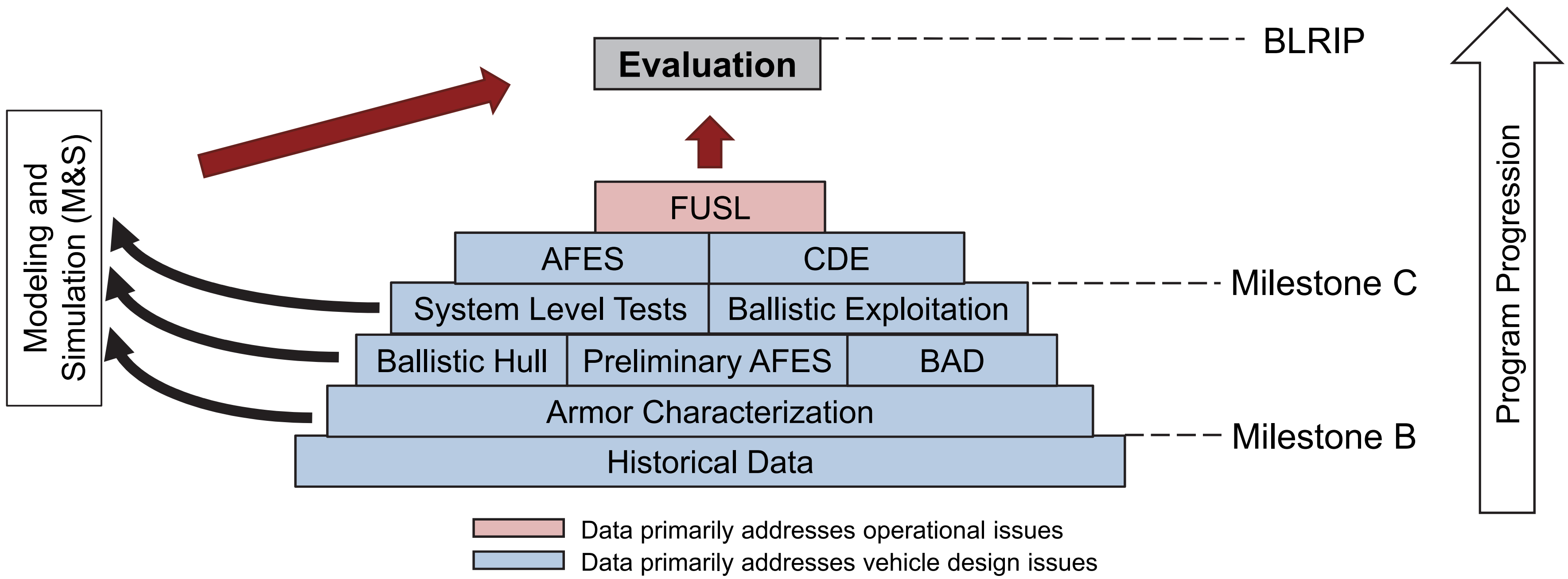
Choose Test Points



Number	Variant	Threat Severity	Shot Line
1	Variant A	Beyond Threshold	Driver
2	Variant B	Beyond Threshold	Engine
3	Variant C	Beyond Threshold	Commander
4	Variant D	Beyond Threshold	Undertrack
5	Variant E	Threshold	Commander
6	Variant A	Threshold	Driver
7	Variant B	Threshold	Engine
8	Variant E	Beyond Threshold	Undertrack
9	Variant D	Beyond Threshold	D Crew
10	Variant C	Beyond Threshold	C Crew
11	Variant A	Beyond Threshold	A Crew
12	Variant B	Beyond Threshold	B Crew
13	Variant E	Beyond Threshold	E Crew
14	Variant D	Beyond Threshold	Commander
15	Variant A	Beyond Threshold	Engine

Acronyms: FUSL – Full-Up System-Level

FUSL is a capstone activity that builds off of knowledge from previous testing and M&S to fill critical knowledge gaps



Acronyms: AFES – Automatic Fire Extinguishing System; BAD – Behind Armor Debris; BLRIP – Beyond Low-Rate Initial Production; CDE – Controlled Damage Experiments; FUSL – Full-Up System-Level; M&S – Modeling and Simulation

FUSL LFT is currently planned by relying on SME input

Current FUSL LFT planning methods rely heavily on SME opinion from several key groups including:

- DOT&E
- PM
- Service evaluators
- User representatives
- Threat community

These groups identify data gaps for FUSL LFT by considering the operational context and expected missions, historical test data, combat data, M&S capabilities, and scope of the planned building block tests

The most critical shots to test in FUSL LFT…

… result in **troop casualties**

… cause loss of **mission essential functions**

… have **uncertainty in test outcome**

Acronyms: LFT – Live Fire Testing; M&S – Modeling and Simulation; PM – Program Manager; SME – Subject Matter Expert

Example: Defining the space of potential shots for a notional armored vehicle

- Consider a notional family of vehicles with five variants
- Some common elements across variants (engine, driver), but crew compartment is unique for each variant
- For simplicity, let's only consider one threat – underbody blast – for this example

Factor	Levels
Variant	Variant A, Variant B, Variant C, Variant D, Variant E
Threat Category	Underbody Blast (UBB)
Threat Severity	Threshold-level, Beyond Threshold-level, Objective-level
Location	Driver, Commander, Crew Compartment, Engine, Undertrack

Full Factorial Design: Up to **75** possible UBB events!

For FUSL, most of the design space is not relevant

Adequate information is available from other sources

Some combinations result in obvious outcomes (severe overmatch or undermatch)

However, there are some shots where our uncertainty of the outcome is relatively high, and our expectation is that the shot could have a substantial effect on the outcome of a mission

To choose important shots, we need to understand for every potential shot:

- How severely do we expect this shot to affect vehicle and crew survivability?
- How certain are we [will we be] pre-FUSL of this estimate?
- How similar is this shot to other potential FUSL shots?

Acronyms: FUSL – Full-Up System-Level

The most critical shots to see in FUSL testing are those that have potentially high, but uncertain mission effects

- The magnitude of a shot's mission effects is based on the troop casualties, mission essential functions, and the operational cost to recover and repair the vehicle
- Most of our FUSL LFT resources should be allocated on shots where we are uncertain about the outcome of the event (i.e., low certainty)

High certainty shots:

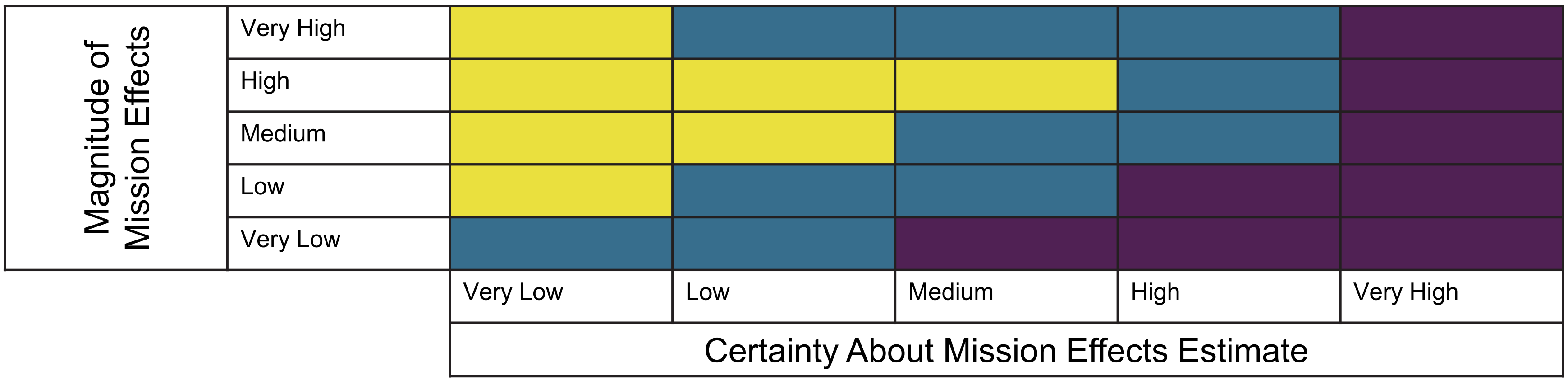
- Similar to shots from similar systems
- Threat is well-understood
- Threat overmatches (or undermatches) system
- Validated, high-quality M&S available
- Pre-FUSL testing will provide substantial relevant information

Low certainty shots:

- Minimal relevant legacy testing
- Novel/untested threat
- Threat level is close to design thresholds
- Existing M&S isn't validated
- Component-level testing doesn't cover synergistic effects from multiple components being hit

Acronyms: FUSL – Full-Up System-Level; LFT – Live Fire Testing; M&S – Modeling and Simulation

We combine our assessment of mission effects and our certainty about these effects to determine the criticality of each shot



Color	Importance
	FUSL shots will address critical knowledge gaps; the vast majority of these knowledge gaps should be answered
	FUSL shots will address moderate knowledge gaps; many of these knowledge gaps should be answered
	FUSL shots will address minor knowledge gaps; few of these knowledge gaps need to be answered

Acronyms: FUSL – Full-Up System-Level

Factors and shot relationships

In DOE, an explicitly or implicitly specified model describes relationships between each design point

- Common factors & interaction effects
- Error distributions
- Hypothesis tests

The FUSL design space doesn't lend itself well to this, so instead we use subjective assessments of pairwise shot "similarity" bounded by a few rules:

- If two shots contain no common factors, they are completely unrelated (similarity = 0)
- If all factors are identical for two shots, they are functionally identical (similarity = 1)
- All factors are not equivalent when it comes to evaluating similarity

Similarity between Shot 1 and Shot 2 will ultimately be used to tell us how much information we learn about FUSL Shot 2 when we observe FUSL Shot 1

Acronyms: DOE – Design of Experiments; FUSL – Full-Up System-Level

Example: How similar are these two shots?

Shot 1

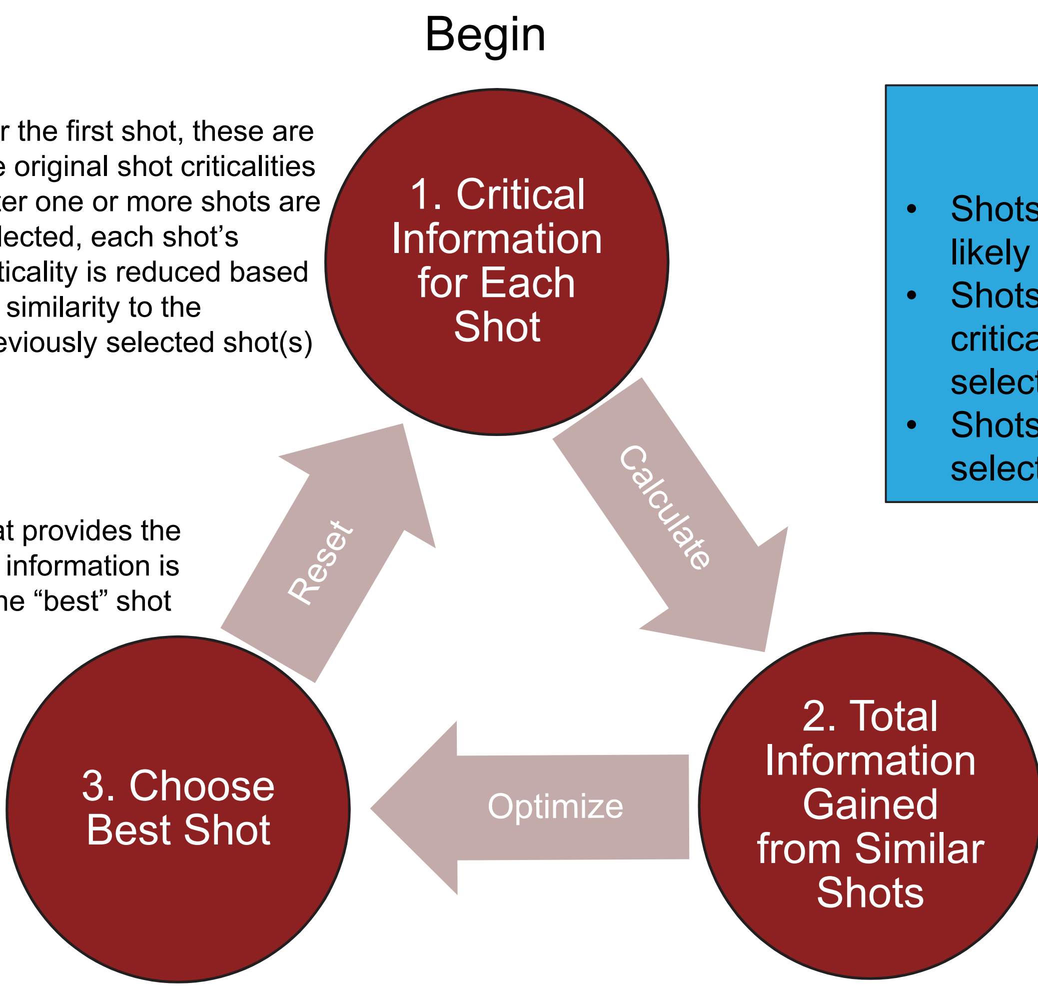
Variant	A
Threat Category	Underbody Blast
Threat Severity	Threshold-level
Location	Under the engine

Shot 2

Variant	B
Threat Category	Underbody Blast
Threat Severity	Objective-level
Location	Under the engine

How much would we learn from doing Shot 2, if we have already conducted Shot 1?

FUSL shot selection algorithm



- For the first shot, these are the original shot criticalities
- After one or more shots are selected, each shot's criticality is reduced based on similarity to the previously selected shot(s)

- Shots with many related shots with high criticality will provide lots of critical information

- The shot that provides the most critical information is chosen as the "best" shot

<u>Shot Prioritization</u>

- Shots with high "criticality" are more likely to be selected
- Shots that are related to many other critical shots are more likely to be selected
- Shots that aren't related to previously selected shots are likely to be selected

Acronyms: FUSL – Full-Up System-Level

The algorithm generates information that helps analysts determine the optimal number of FUSL tests

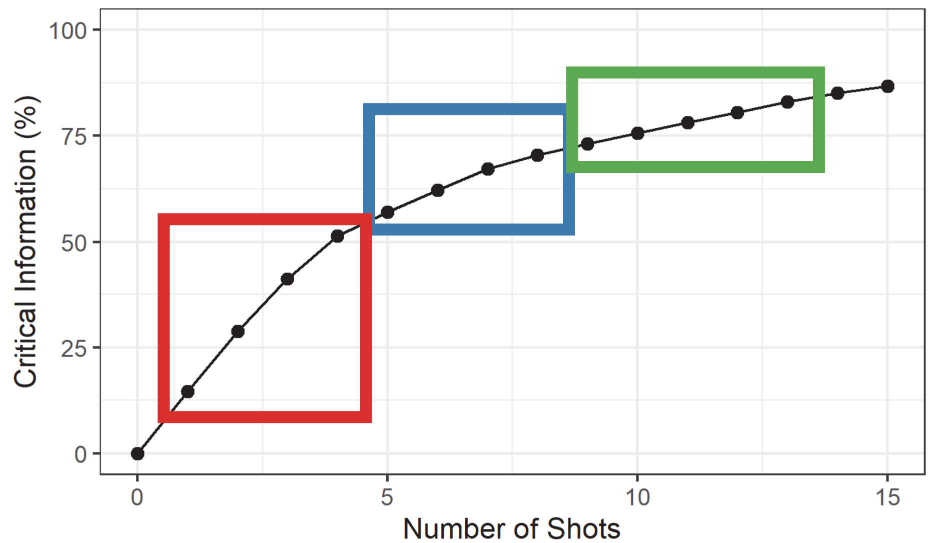
First 15 Shots (in order)

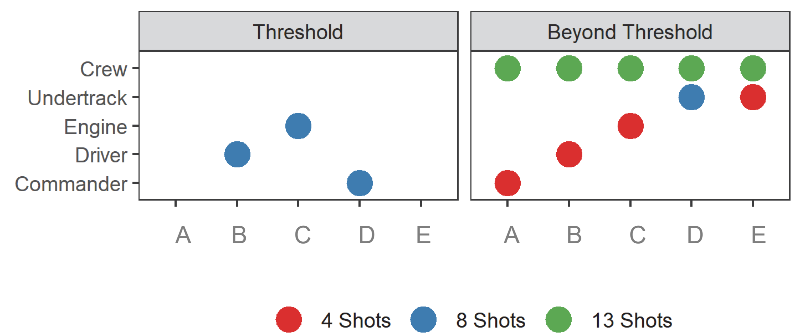
Number	Variant	Threat Severity	Shot Line
1	Variant A	Beyond Threshold	Driver
2	Variant B	Beyond Threshold	Engine
3	Variant C	Beyond Threshold	Commander
4	Variant D	Beyond Threshold	Undertrack
5	Variant E	Threshold	Commander
6	Variant A	Threshold	Driver
7	Variant B	Threshold	Engine
8	Variant E	Beyond Threshold	Undertrack
9	Variant D	Beyond Threshold	D Crew
10	Variant C	Beyond Threshold	C Crew
11	Variant A	Beyond Threshold	A Crew
12	Variant B	Beyond Threshold	B Crew
13	Variant E	Beyond Threshold	E Crew
14	Variant D	Beyond Threshold	Commander
15	Variant A	Beyond Threshold	Engine

Highly uncertain shots generalizable across variants

Threshold-level threats

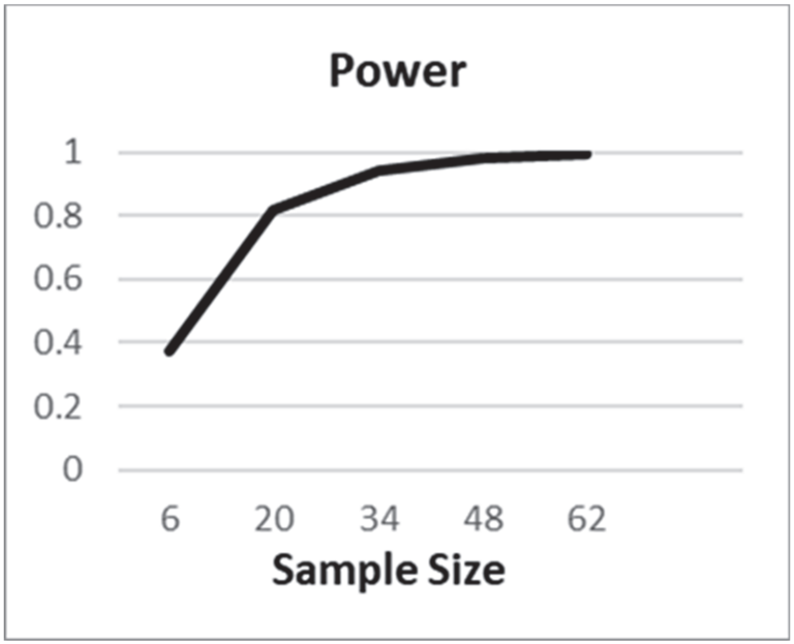
Important but unique shotlines (low similarity)



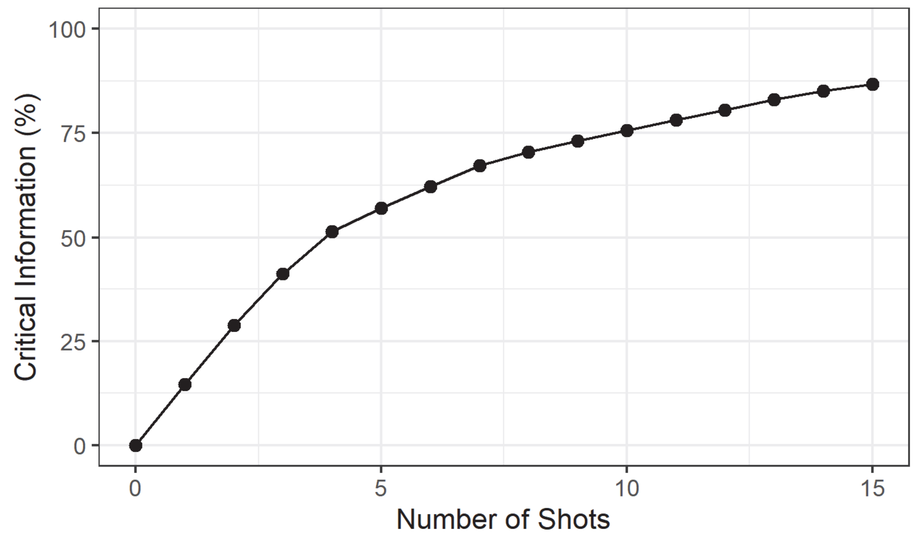


The percentage of critical information gained can be thought of analogously to a power curve from DOE

DOE for OT



DOE for FUSL LFT



Acronyms: DOE – Design of Experiments; FUSL – Full-Up System-Level; LFT – Live Fire Testing; OT – Operational Testing

Summary

- We developed a quantitative optimization algorithm for determining FUSL test scope that formalizes the FUSL test planning process
  - This approach is analogous to the type of DOE used in OT&E
- This approach can be used to help analysts select the optimal size of FUSL tests
  - Future work could expand this work to aid analysts select specific shotlines for FUSL LFT
- As a proof of concept, we have illustrated this technique on a notional armored vehicle
- Future work including sensitivity studies, algorithm improvements, and retrospective studies on programs of record should be completed prior to implementation

Acronyms: DOE – Design of Experiments; FUSL – Full-Up System-Level; OT&E – Operational Test and Evaluation

Questions?

Backup

Notional Scoring Rubrics

Criticality of Mission Effects

Score	Description	Example
1	Very low mission effects	Engine remains running; no injuries to occupants; some damage to non-critical components; FMC
2	Low mission effects	Damage to low importance (but still critical) components; no occupant injuries
3	Medium mission effects	Damage to at least one (high value) critical component or multiple low value critical components, OR moderate (AIS 2) occupant injuries to a single occupant
4	High mission effects	Serious injuries (AIS 3) injuries to a single occupant or AIS 2+ injuries to multiple occupants, OR significant loss of mobility or firepower, but not both
5	Very high mission effects	Severe injuries (AIS 4+) to a single occupant or AIS 3+ injuries to multiple occupants, OR catastrophic kill of vehicle, OR both firepower and mobility are significantly impaired (near zero functionality)

Certainty

Score	Description	Example
1	Very low certainty	Essentially a random guess; no historical data with similar threats or vehicles; no M&S; no relevant early building block testing
2	Low certainty	Some historical data or early testing, but unknown synergistic effects at the vehicle level; no M&S
3	Medium certainty	Either M&S or historical data plus component level testing, but not both (i.e., no validation data for M&S); some limitations of available data
4	High certainty	M&S with either historical data or early component testing to help validate; some variation in mission effects; minor limitations of data sources
5	Very high certainty	Practically a known outcome; at least 2 different data sources (M&S, historical data, early component tests) with little variation in mission effects; severe overmatch or undermatch

Acronyms: AIS – Abbreviated Injury Scale; FMC – Fully Mission Capable; M&S – Modeling and Simulation

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION

1. REPORT DATE 09-03-2022	**2. REPORT TYPE** Draft Final	**3. DATES COVERED** **START DATE**	**END DATE** Mar 2022

4. TITLE AND SUBTITLE
DATAWorks 2022: A New Method for Planning Full-Up System-Level (FUSL) Live Fire Tests

5a. CONTRACT NUMBER Separate Contract	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
5d. PROJECT NUMBER C9107	**5e. TASK NUMBER** C9107	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Butler, Lindsey, D.; Avery, Matthew, R.; Gardner, William, G.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 730 East Glebe Road Alexandria, Virginia 22305	**8. PERFORMING ORGANIZATION REPORT NUMBER** NS D-32966 H 2022-000043

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**	**11. SPONSOR/MONITOR'S REPORT NUMBER**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Public release approved. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
Project Leader: Couch, Mark A.

14. ABSTRACT
Planning Full-Up System-Level (FUSL) Live Fire tests is a complex process that has historically relied solely on subject matter expertise. In particular, there is no established method to determine the appropriate number of FUSL tests necessary for a given program. We developed a novel method that is analogous to the Design of Experiments process that is used to determine the scope of Operational Test events. Our proposed methodology first requires subject matter experts (SMEs) to define all potential FUSL shots. For each potential shot, SMEs estimate the severity of that shot, the uncertainty of that severity estimate, and the similarity of that shot to all other potential shots. We developed a numerical optimization algorithm that uses the SME inputs to generate a prioritized list of FUSL events and a corresponding plot of the total information gained with each successive shot. Together, these outputs can help analysts determine the adequate number of FUSL tests for a given program. We illustrate this process with an example on a notional ground vehicle. Future work is necessary prior to implementation on a program of record.

15. SUBJECT TERMS
Live Fire Testing; Design of Experiments (DOE); Full-Up System Level (FUSL); Test Planning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	SAR	29

19a. NAME OF RESPONSIBLE PERSON Mark A. Couch	**19b. PHONE NUMBER** 703-845-2530