INSTITUTE FOR DEFENSE ANALYSES

Introduction to Human-Systems Interaction in Operational Test and Evaluation Course

Dr. Keyla Pagan-Rivera, Project Leader

Dr. Adam M. Miller

OED Draft

June 2024

IDA Product ID 3002009

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, Task BD-9-229998, "TestSci Training," 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 Dr. V. Bram Lillard from the Science, Systems and Sustainment Division and consisted of Dr. Gillen B. Brown, Dr. John Haman, and Dr. Miriam Armstrong from the Operational Evaluation Division, and Dr. Megan Gelsinger from the Science, Systems and Sustainment Division.

For more information:
Dr. Keyla Pagan-Rivera, Project Leader
kpaganri@ida.org • (703) 845-6936

Dr. Heather M. Wojton, Director, Operational Evaluation Division
hwojton@ida.org • (703) 845-6811

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Product ID 3002009

Introduction to Human-Systems Interaction in Operational Test and Evaluation Course

Dr. Keyla Pagan-Rivera, Project Leader

Dr. Adam M. Miller

This page intentionally left blank

Executive Summary

The purpose of this training is to introduce operational test and evaluation (OT&E) practitioners to the relevant policies, concepts, and methods for conducting human-systems interaction (HSI) evaluations. This training expands upon previous HSI trainings offered at IDA by providing a theoretical framework for understanding how HSI supports system performance. The training explores this framework over the entire HSI evaluation timeline—from conceptualizing a test to reporting an evaluation—to show how HSI principles contribute to system evaluations.

The goal of an operational HSI evaluation is to characterize how HSI affects the user's ability to accomplish their operational mission. To do this, testers must define their operational research questions in terms of user behavior (i.e., tasks performed by users in support of system performance), collect valid user data under operational conditions, and then explain the relationship between user behavior and system performance across the operational envelope.

HSI is described in terms of HSI concepts, such as usability and workload, which explain how the relationship between users, systems, and the environment affects user behavior. Evaluators collect data on HSI concepts by administering surveys, conducting interviews, or directly measuring user behavior. The strongest evaluations combine all three measurement types to 'triangulate' HSI concepts. Measures should be administered during or shortly after task performance, and at a sufficient tempo to collect data across the operational envelope, all while taking care to avoid compromising operational realism.

HSI data should be analyzed in terms of the relationship between user behavior and system performance. Pertinent questions include how operational conditions affect HSI, how HSI affects task performance, and how task performance impacts overall system performance. Quantitative data should be modeled in terms of operational factors, while qualitative data should be organized according to consistent, operationally relevant themes.

Lastly, evaluating and reporting HSI data requires combining the behavioral, survey, and interview analyses into a coherent overall evaluation of how HSI affects system performance across the operational envelope. Traditionally, HSI evaluations are reported in terms of suitability, but they may also appear in effectiveness or survivability sections.

A. Introduction to HSI

DoD 5000.95 instructs the T&E community to evaluate HSI in an operational environment. To do this, testers must characterize how operational factors affect the ability of users to interact with the system under test. This requires measuring user behavior in operational settings. It also requires characterizing how HSI supports system performance across the operational envelope.

B. The system from an HSI perspective

Users interact with systems via system components; semi-independent subsystems that may be designed with a user interface. For example, a car contains a 'driving component' that includes the steering wheel, acceleration, and brake pedals (Figure 1).

System components facilitate user tasks that support overall system performance, such as driving. HSI evaluations should characterize the ability of a user to perform a system-relevant task using a system component. HSI evaluations should also consider how task performance supports overall system performance.

Users, components, and tasks are all empirically defined and context-dependent. Therefore, it is essential to observe users interacting with the system within a mission context. By observing users, researchers can see which system components they interact with, what tasks they accomplish using system components, and how these tasks contribute to overall system performance.

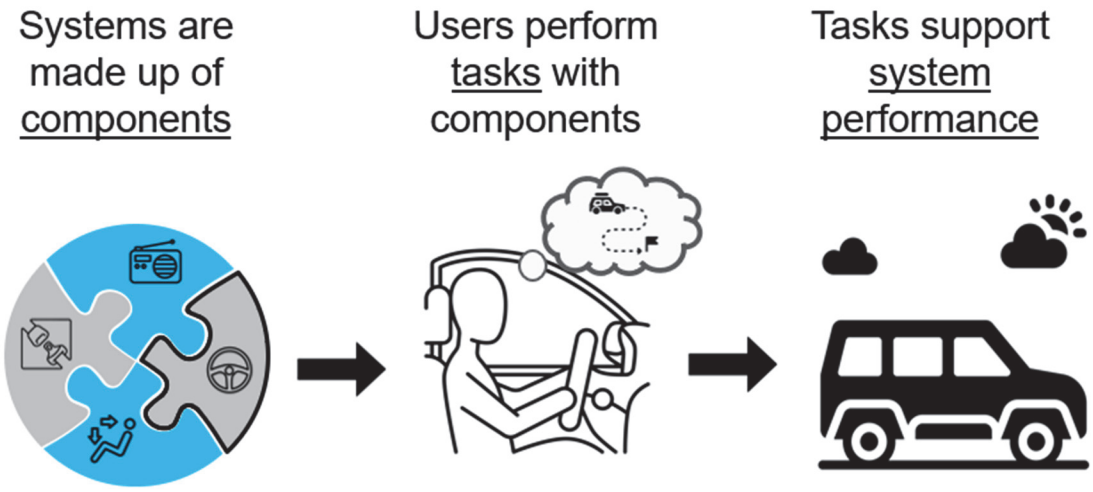


Figure 1. Users support system performance by performing tasks on system components.

C. HSI concepts

DOT&E's 2019 HSI guidance memo directs the community to characterize system users, system components, and system tasks in terms of HSI concepts, which describe HSI in measurable terms like training, usability, and workload (Figure 2).

Most systems are deployed with an accompanying training program, often called "new equipment training." HSI evaluations should evaluate whether the new equipment training effectively prepares users to employ the system to complete their operational mission.

HSI evaluations should also characterize the human factors engineering of system components that support task performance. In particular, HSI evaluations should consider the usability of system components, which refers to how intuitive, effective, and frustration-free the interfaces are.

When characterizing the tasks that users perform on system components, HSI evaluations should consider the user's mental state while performing the task, including their available mental resources (workload), their beliefs about the system (e.g., trust), and their beliefs about the environment (e.g., situational awareness).

Other relevant HSI concepts relate to ergonomics and user demographics, such as whether the system accommodates the user's physique, or how the user experience level affects task performance.

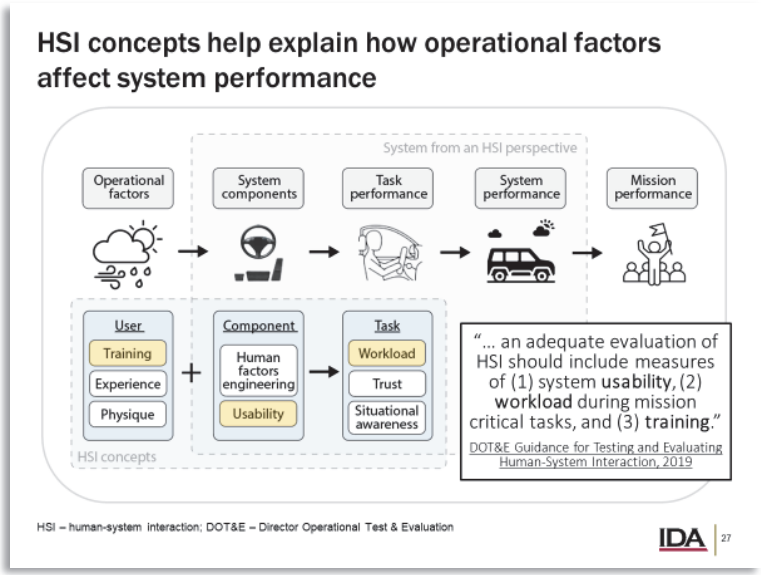


Figure 2: HSI concepts support a characterization of system performance.

D. HSI measures

HSI measures relevant to T&E fall within three major categories: behavior, surveys, and interviews. The strongest HSI evaluations combine evidence from all three categories to 'triangulate' HSI concepts, increasing the specificity of our conclusions about how HSI concepts affect system performance.

Behavioral measures are used to evaluate aspects of task performance, such as timeliness and accuracy. Measuring task performance is crucial to understanding HSI

in the system because task performance is the main way that users contribute to system performance.

Self-report measures such as surveys and interviews elucidate the relationship between HSI concepts and task performance. These measures ask users directly about their experiences while performing the task. Validated surveys are particularly useful for efficiently quantifying HSI concepts. Interviews and focus groups are flexible methods for collecting in-depth information from users with rare and valuable experiences.

E. HSI design of experiments

System performance often varies across operational conditions due, in part, to how those conditions affect HSI. For example, a user that experiences moderate workload while driving in the light might experience high workload in the dark, which could degrade system performance specifically at nighttime.

To characterize HSI across operational conditions, testers should employ a design of experiments (DOE) approach that varies operational factors during an operational test. The HSI evaluation will often use the same DOE as the overall system so that evaluators can test how HSI contributes to the relationship between operational factors and system performance.

The DOE should be accompanied by an administration plan detailing what HSI measures will be administered, to whom they will be administered, and when they will be administered during the test. Ideally, testers would administer behavioral measures, surveys, and interviews during every mission (Figure 3). However, testers must also ensure that the data collection procedures don't compromise the operational realism of the test. One approach is to measure task performance during every mission with automated instruments that are invisible to the user, along with administering validated surveys after every mission and holding focus groups at the end of each day.

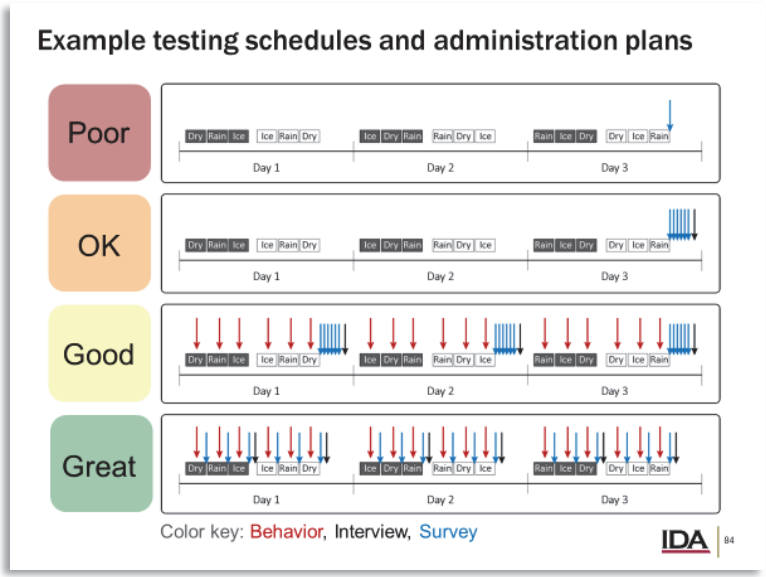


Figure 3: Administer HSI measurements to support conclusions about the impact of operational factors.

F. Analyzing HSI data

The primary goal of HSI data analysis is to characterize the relationships between HSI concepts and operational factors. Building evidence for a relationship between HSI concepts and operational factors starts with characterizing variance in the HSI data according to the DOE and administration plan. Data from quantitative measures, including behavioral data and survey responses, should be modeled as a function of the operational factors in the DOE (Figure 4). Data from qualitative measures, including interviews and open-ended survey questions, should be reviewed to identify common themes that describe the experience of the user while performing tasks, and that relate the user's experiences to operational factors. Testers should also make special note of user feedback that identifies unexpected HSI concerns.

G. HSI evaluation and reporting

The goal of the HSI evaluation is to determine how HSI affects the ability of a unit equipped with the system to accomplish its combat mission. This is generally accomplished by characterizing how HSI concepts affect task performance, and how task performance affects system performance.

The HSI evaluation is the culmination of the DOE and triangulation approaches to HSI data collection. Evaluators can make a compelling case about how HSI affects system performance by synthesizing what they learn from analyzing the behavioral, survey, and interview data. Evaluations should start by highlighting the most relevant analyses to each question at hand, such as by describing driving performance in terms of instrumented behavioral data.

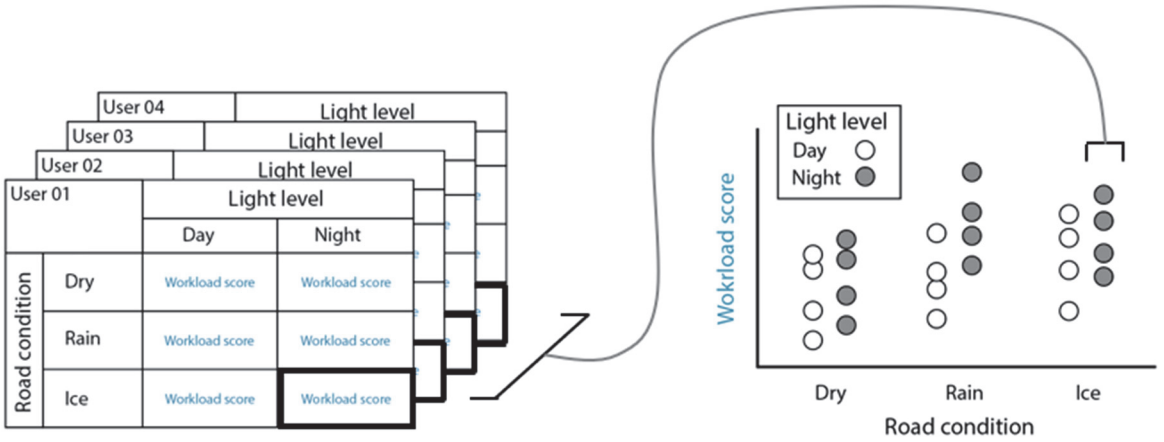


Figure 4. Analyze HSI data in terms of operational factors.

However, evaluations should then synthesize across analyses to characterize how HSI affects system performance more broadly. For example, if (1) the analysis of the behavioral data indicates that drivers had difficulty staying in their lane at night, (2) the analysis of the survey data shows that driver workload was higher at night, and (3) drivers reported during interviews having difficulty managing their task load while driving at night, then an analyst might synthesize these three analyses to conclude that driving performance was worse at night due to increased workload caused by a burdensome task load.

Lastly, reporting the HSI evaluation requires communicating HSI conclusions within the larger effectiveness, suitability, and survivability evaluation. Traditionally, HSI evaluations are reported in terms of suitability. Occasionally, however, some elements of an HSI evaluation may belong in the effectiveness or survivability sections, particularly when user task performance directly affects these areas. In all cases, analysts should prioritize communicating how HSI affects system performance as opposed to simply describing the state of HSI concepts.

H. Reviewing test plans

When reviewing test plans, reviewers should consider whether the plan—if executed as planned and with the current level of specificity—will enable an analyst to draw conclusions about the relationships between operational factors, HSI concepts, task performance, and system performance.

In particular, reviewers must consider whether the test plan conforms to DOT&E's 2019 Guidance for Testing and Evaluating HSI. The guidance instructs testers to describe how they will collect HSI data across all relevant operational conditions. The guidance also instructs testers to include a data collection plan that summarizes what HSI concepts they will measure, what methods they will use to measure them, and when and how they will collect the data.

Additionally, the guidance notes the importance of measuring HSI concepts with quantitative and qualitative measurements that are linked to a DOE to evaluate how HSI relates to measures of operational performance. Reviewers should also consider whether the HSI concepts of interest will be sufficiently captured by the proposed HSI measures, including whether the plan will collect data on the most relevant users, components, and tasks.

Introduction to Human-Systems Interaction in Operational Test and Evaluation

Behavior = *f*(Person, Environment)

(Lewin, 1951)

Adam M. Miller

Institute for Defense Analyses

730 East Glebe Road ● Alexandria, Virginia 22305

Contents

- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

Contents

What is Human-Systems Interaction (HSI)?

Human-Systems Interaction is the study of interfaces between humans and technical systems

Human Factors Engineering is the application of human principles to product design

Human-Systems Integration is an interdisciplinary approach to understanding and improving the interfaces between humans and technical systems

DOT&E approves test plans and evaluates the adequacy of testing – *this includes HSI!*

DoD Instruction 5000.95

Human Systems Integration in Defense Acquisition

Section 2: Responsibilities

2.4. DIRECTOR OF OPERATIONAL TEST AND EVALUATION (DOT&E).

In addition to the responsibilities in Paragraph 2.6., the DOT&E:

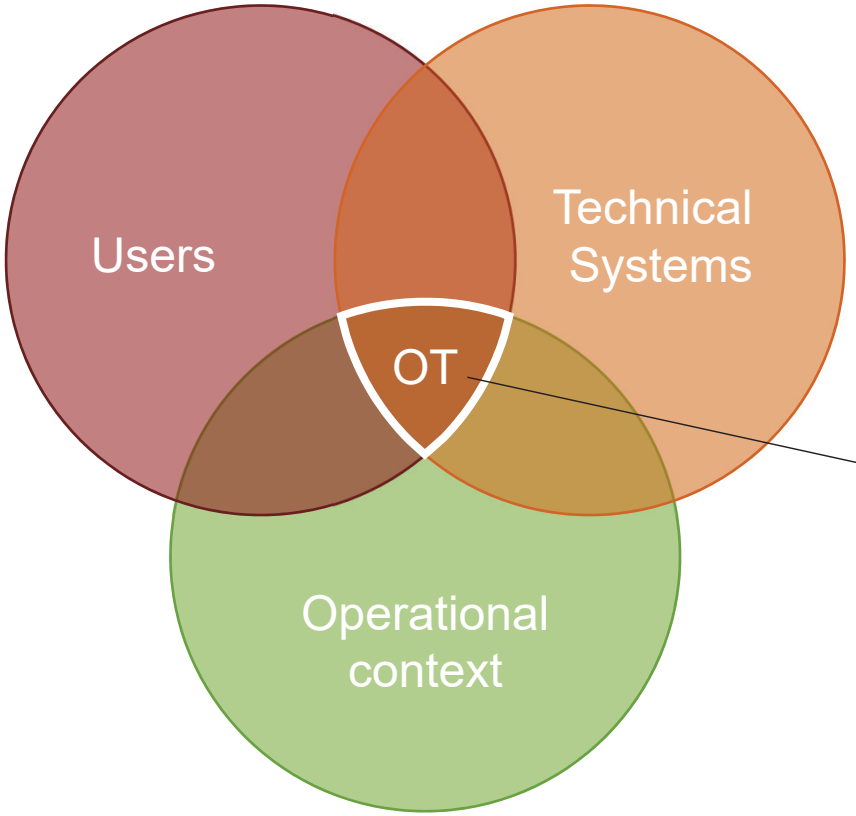
a. Incorporates the adequate evaluation of human systems integration in an operational environment in the DOT&E policy and procedures in accordance with DoDD 5141.02 and this issuance.

b. For programs designated with DOT&E oversight, approves human systems integration plans for usability and user testing for human to machine interface analysis under operational conditions.

HSI – Human-Systems Interaction; DoDD – Department of Defense Directive; DOT&E – Director, Operational Testing and Evaluation

Humans are essential to operations; HSI is essential to operational testing

- Operational Testing (OT): systems in context with users in context



HSI characterizes how system performance is impacted by users, within an operational context

HSI – Human-Systems Interaction; OT – Operational Testing

HSI evaluates systems by measuring users

- Don’t blame the user if the system doesn’t work!
- Technology exists to make users more capable
- Well-designed systems meet users where they are
  - Accommodate human abilities
  - Anticipate, minimize, tolerate error
  - Promote user safety

The Masochist’s Coffeepot illustrates how design contributes to accidents

HSI – Human-Systems Interaction

HSI affects effectiveness, suitability, and survivability

<u>Effectiveness</u> of the Joint Air-to-Ground Missile

"Effectiveness is reduced under **high pilot workloads**…due to…a **cumbersome pilot-vehicle interface** (PVI)."

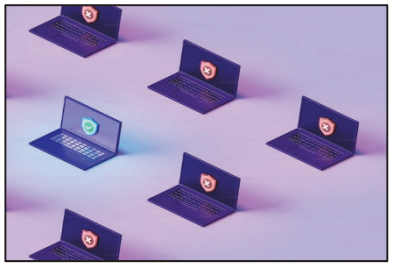
- 2022 FY DOT&E Annual Report, p 115

<u>Suitability</u> of the F-16 Radar Modernization Program

"Pilots are generally satisfied with the **human-systems interface**, although some limitations … resulted in **increased pilot workload** for some tasks."

- 2022 FY DOT&E Annual Report, p 270

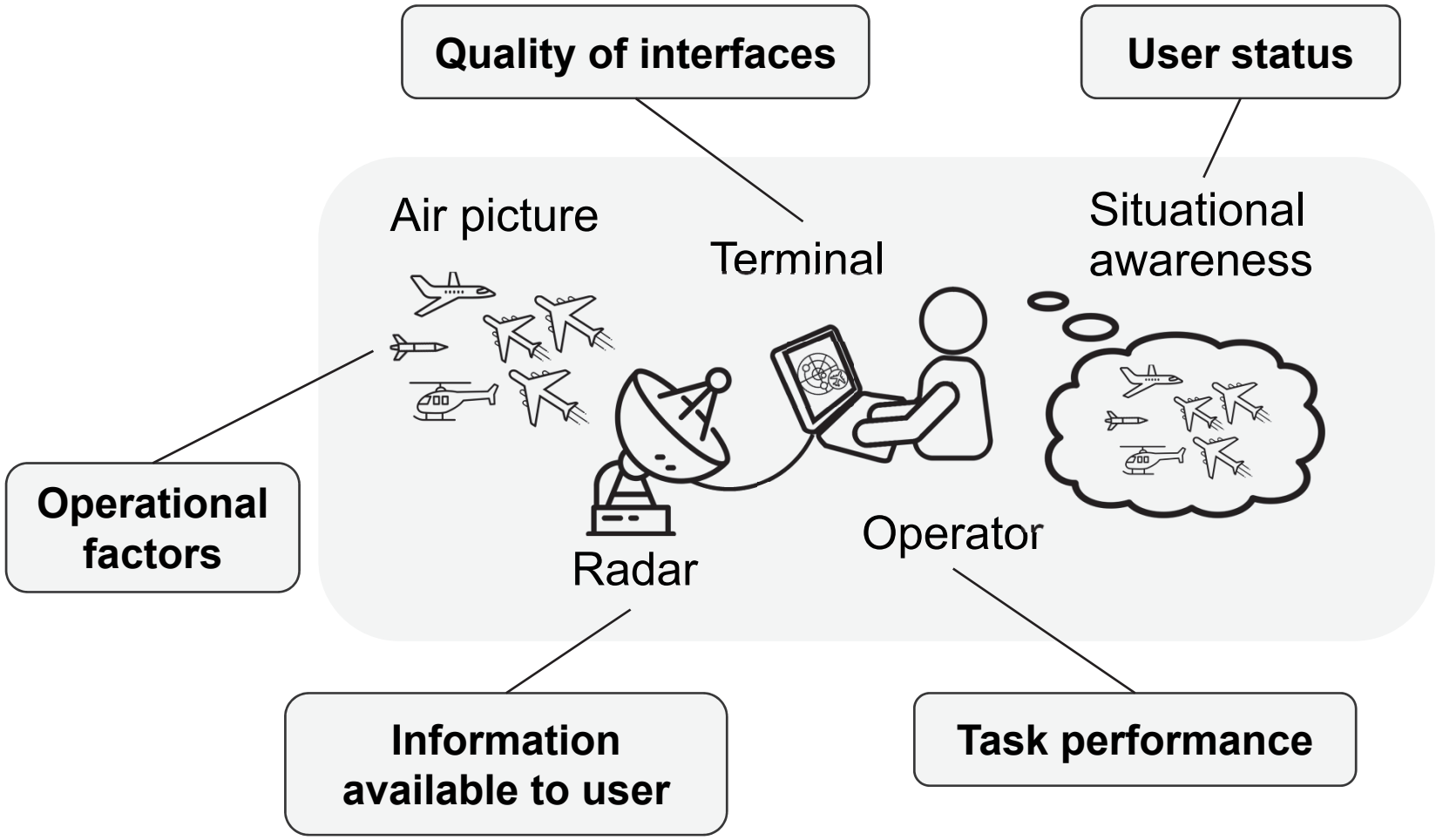
Tradeoffs Between Usability and Cyber <u>Survivability</u>



Tradeoffs between (cyber)security and usability of systems mean that **poor HSI may introduce vulnerabilities** that can affect **operational survivability**.

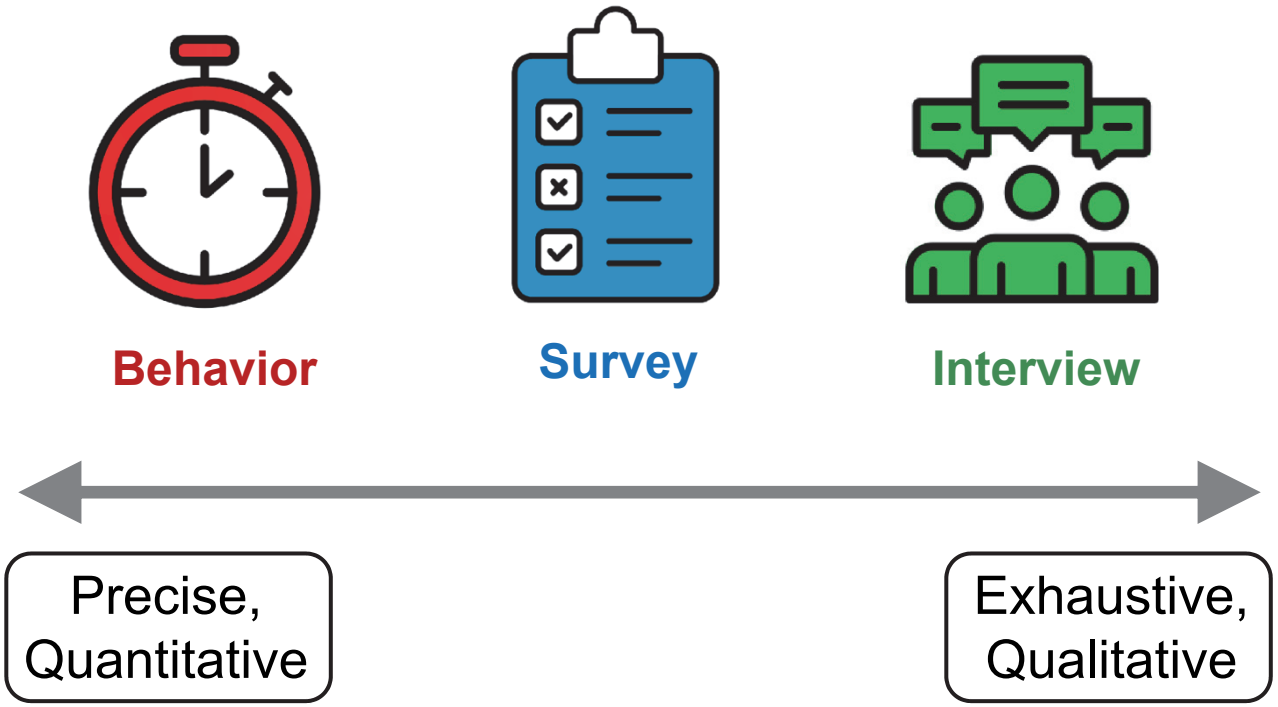
DOT&E – Director, Operational Test and Evaluation; FY – Fiscal Year; HSI – Human-Systems Interaction; PVI – Pilot-Vehicle Interface

Today, we will learn what needs to be measured to support an HSI evaluation



HSI – Human-Systems Interaction

...we will learn about HSI measurement methods



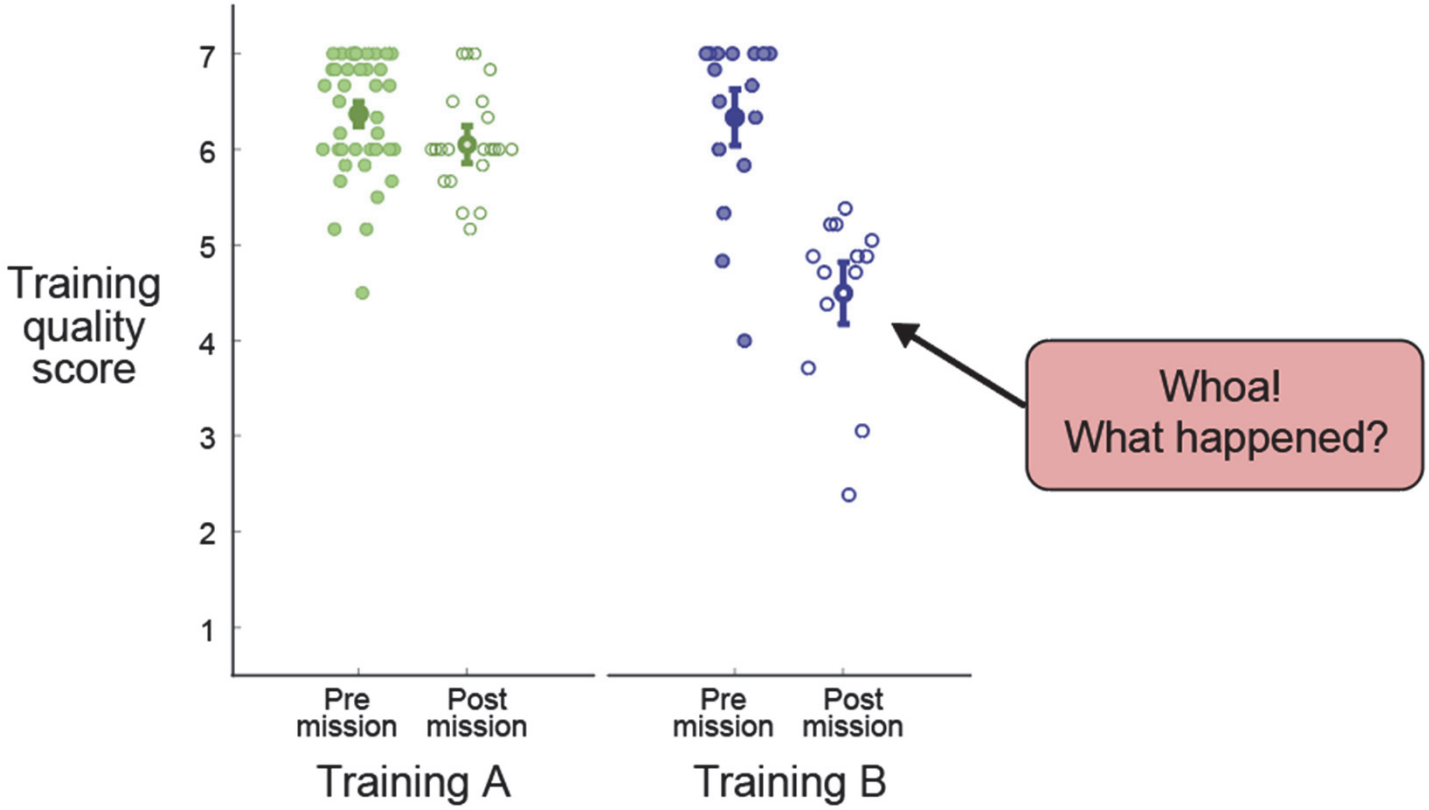
HSI – Human-Systems Interaction

...we will learn about design of experiments for HSI

Road surface conditions	Light level: Day	Light level: Night
Dry	Behavior, Interview, Survey	Behavior, Interview, Survey
Rain	Behavior, Interview, Survey	Behavior, Interview, Survey
Ice	Behavior, Interview, Survey	Behavior, Interview, Survey

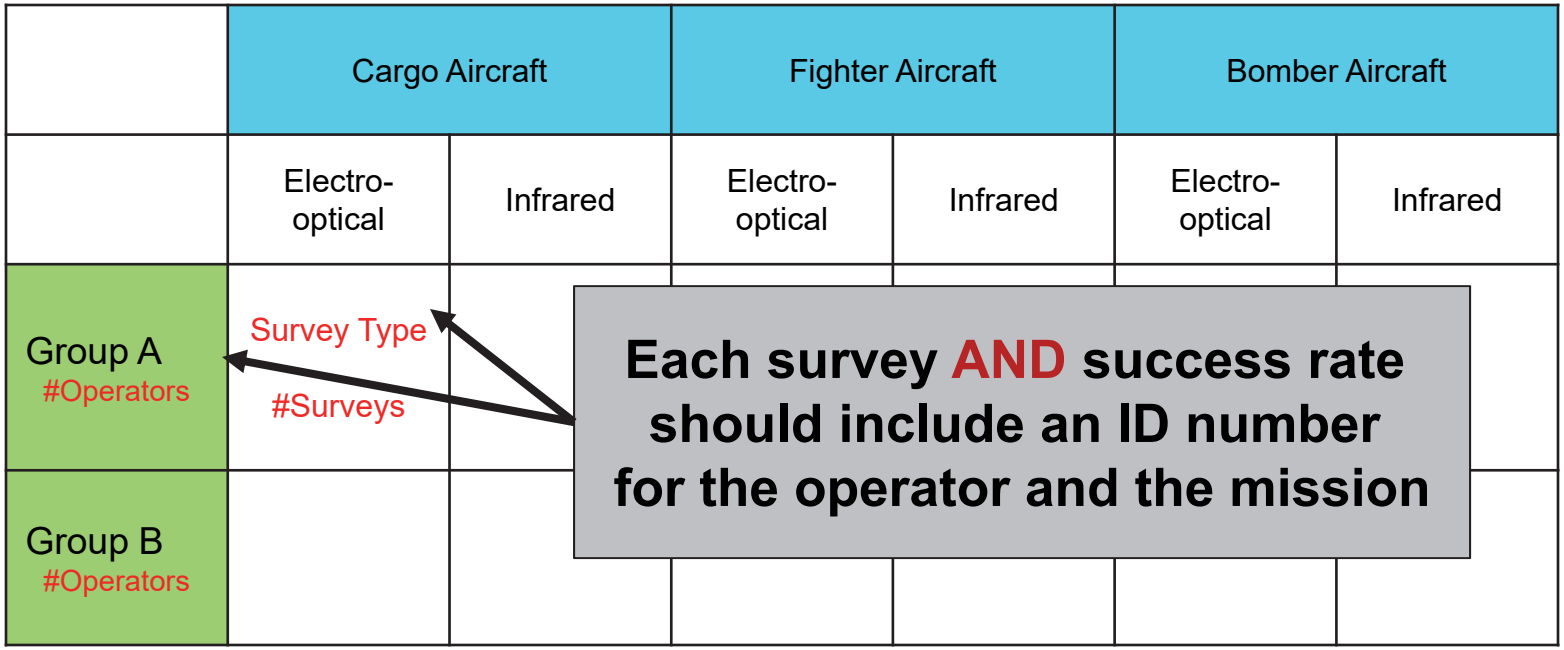
HSI – Human-Systems Interaction

...we will learn how to analyze and report HSI data



HSI – Human-Systems Interaction

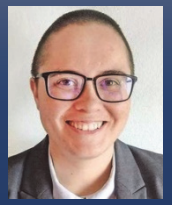
...and we will learn how to review HSI test plans



HSI – Human-Systems Interaction

The Test Science HSI team at IDA is here to help

Air Warfare



Dr. Miriam Armstrong
Human Factors

PhD Human Factors Psych.
Texas Tech Univ.
@ IDA since 2022

Land Warfare

Dr. Elizabeth Green
Human Factors
Test Design

PhD Experimental Psychology
Texas Tech Univ.
@ IDA since 2021

Dr. Adam Miller
Behavioral Science
Data Analysis

PhD Behavioral Neuroscience
Cornell Univ.
@ IDA since 2022

Naval Warfare

Net Centric

Dr. Brian Vickers
AI&A, HSI
Data Science

PhD Cognitive Neuroscience
Univ Michigan, Ann Arbor
@ IDA since 2019

Dr. Sarah Shaffer
Human Factors

PhD Experimental Psych.
Florida International University
@ IDA since 2022

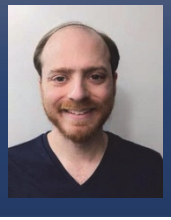
Additional expertise at IDA:

Dr. Dan Porter
AI&A
Human Factors

PhD Cognitive Neuroscience
Univ. of Michigan
@ IDA since 2017

Ms. Rachel Haga
CDAO Lead,
AI&A, Human Factors

MS Aerospace Engineering
Georgia Tech
@ IDA since 2020



Mr. Yosef Razin
Human Factors,
Robotics, Trust

BS Mechanical Engineering
Georgia Tech
@ IDA since 2022

AI&A – Artificial Intelligence and Automation; CAP – Cyber Assessment Program; CDAO – Chief Digital and Artificial Intelligence Office; HSI – Human-Systems Interaction

Contents

The system from an HSI perspective

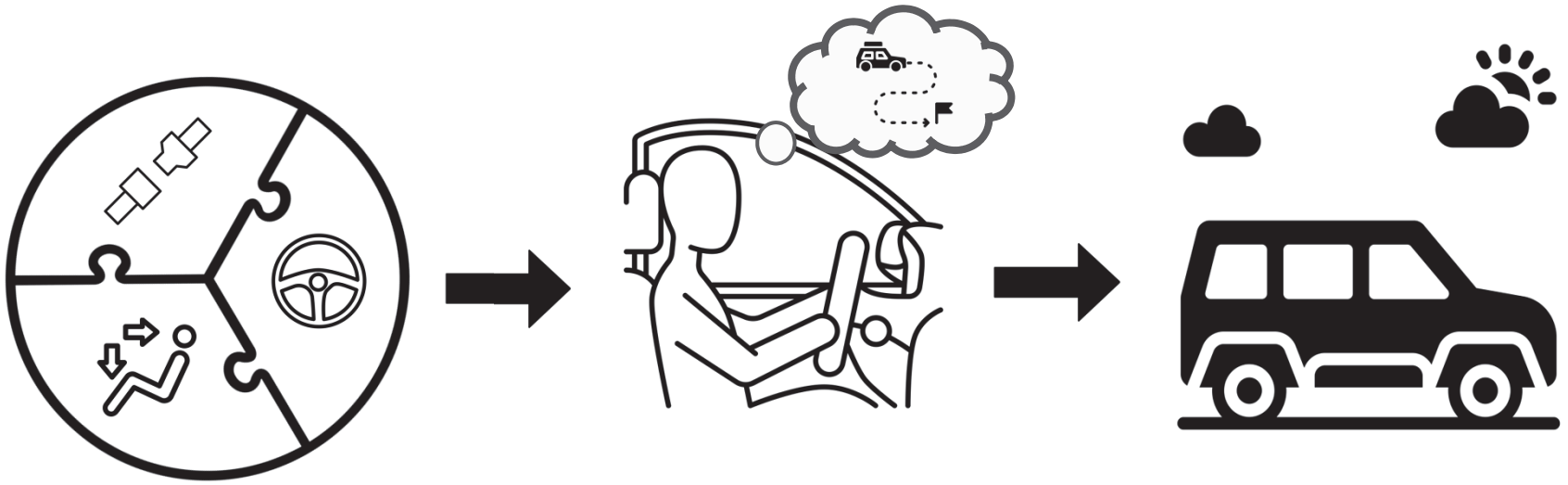
HSI – Human-Systems Interaction

The system from an HSI perspective

Systems are made up of components

Users perform tasks with components

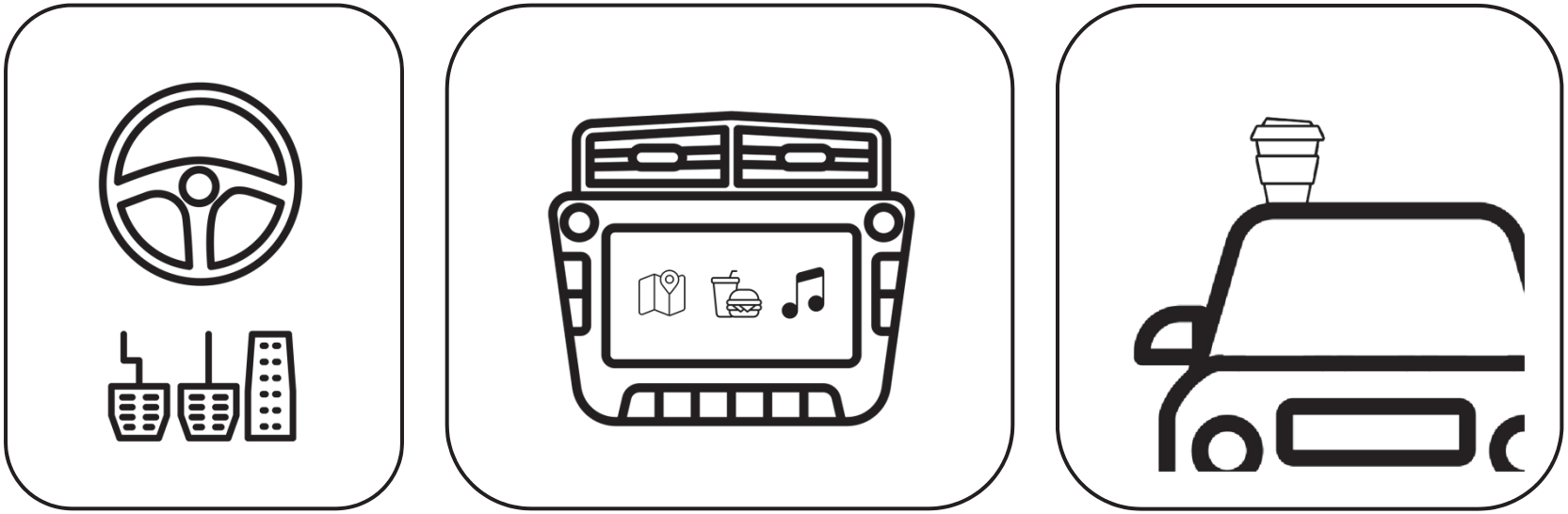
Tasks support system performance



HSI – Human-Systems Interaction

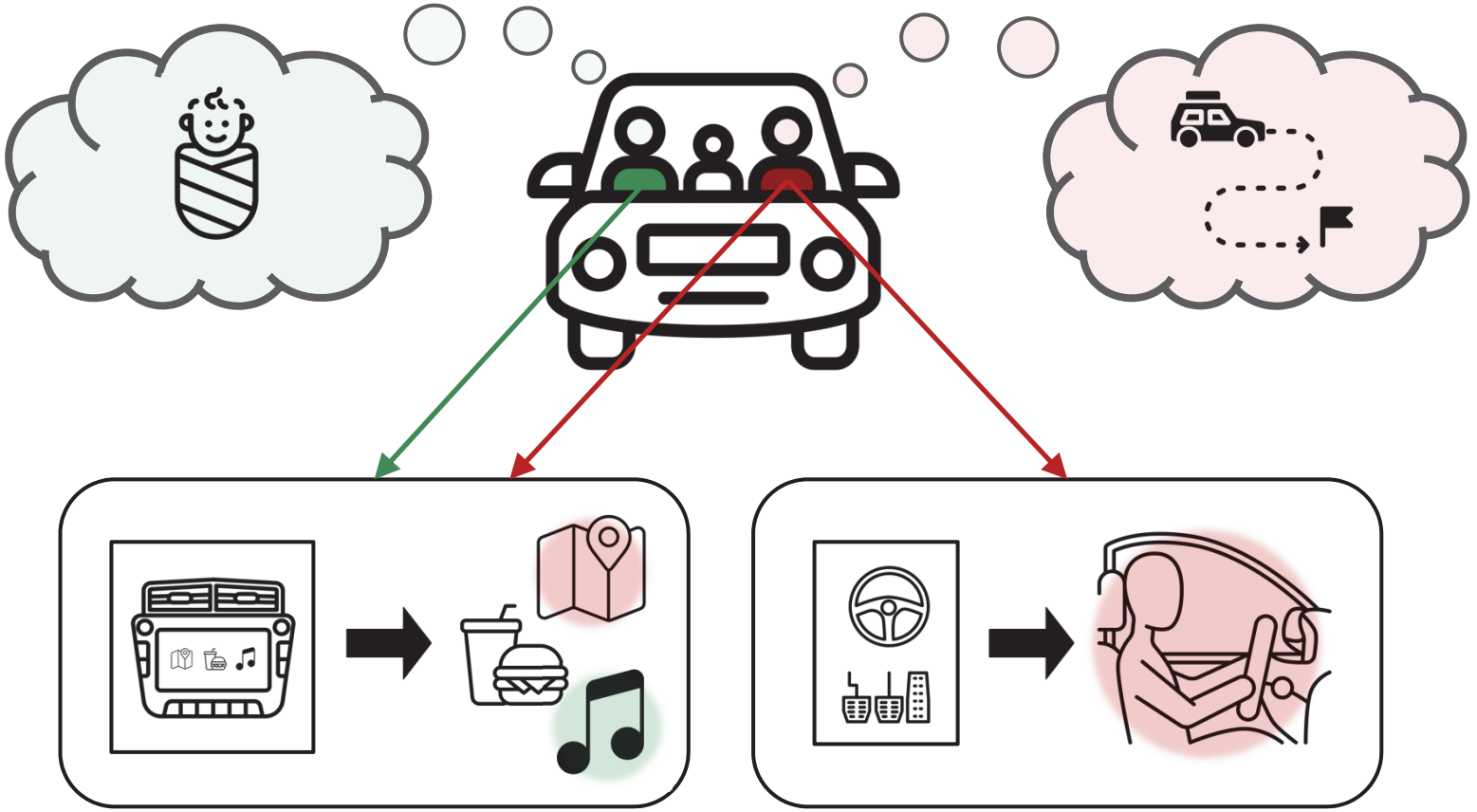
System components

What components of the system do users interact with?



Users perform tasks with system components

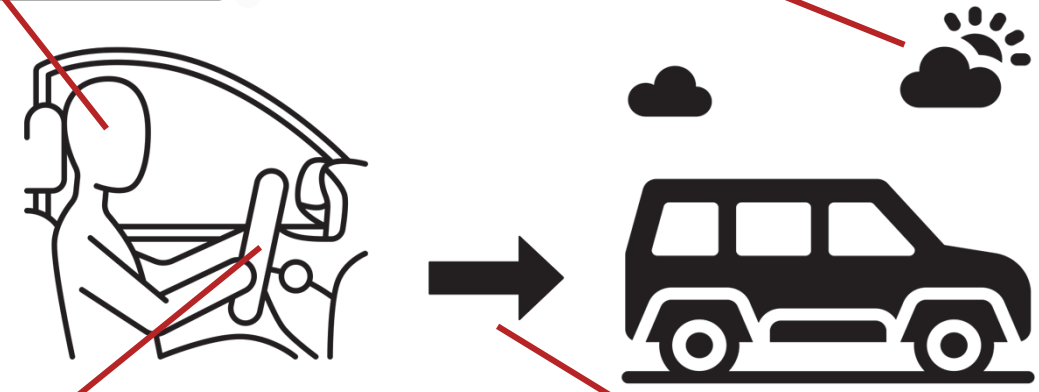
Individual users can perform tasks with multiple components
Individual components can be used for multiple tasks



Task performance supports system performance

How do **user** characteristics affect task performance?

How do **operational factors** affect task performance?

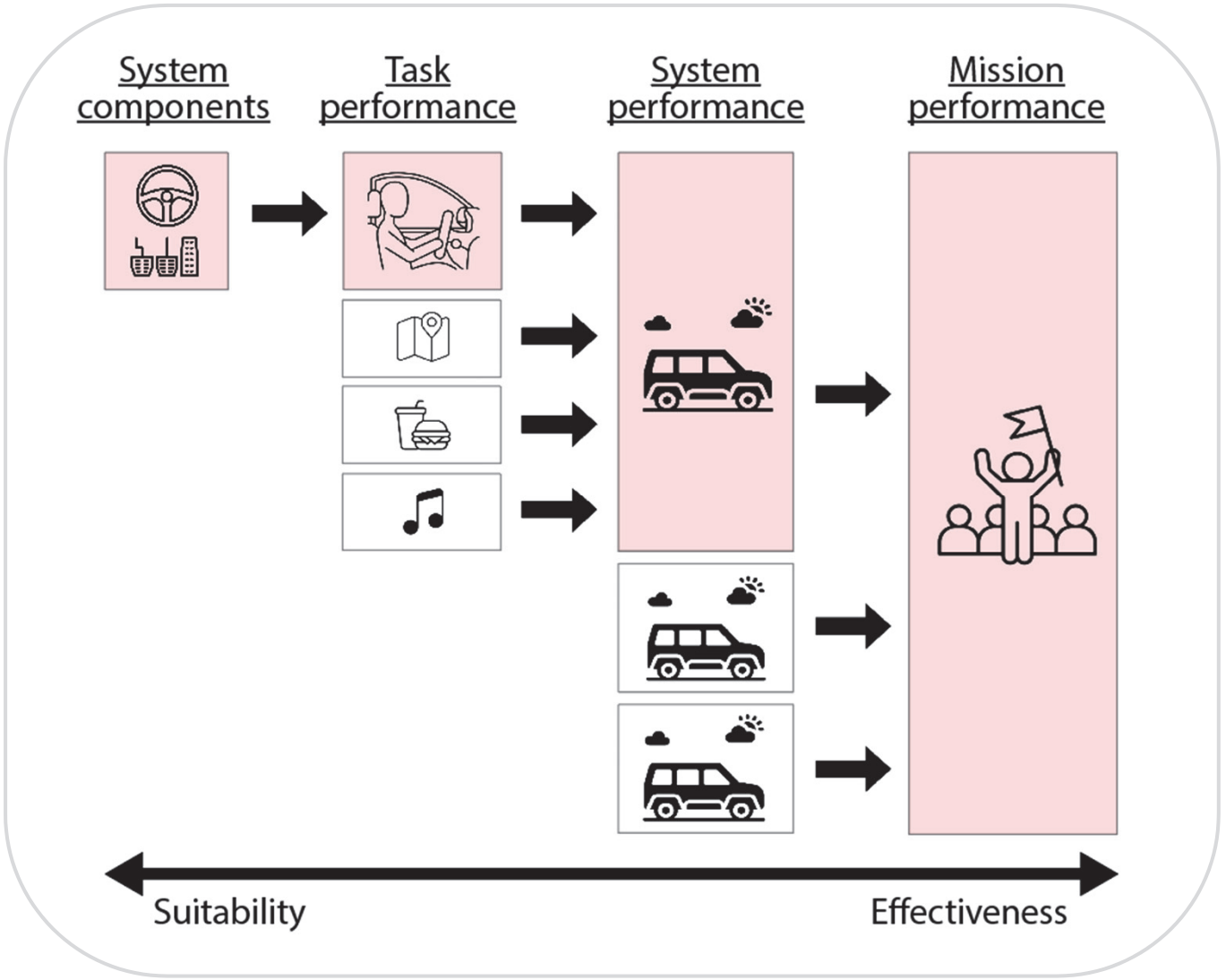


How do **component** characteristics affect task performance?

How does task performance contribute to overall system performance?

HSI – Human-Systems Interaction

Can a unit equipped with the system accomplish its combat mission?



The system from an HSI perspective, key take-aways

- The system: users performing tasks with components
- Let users show you how they use the system
- HSI explores the relationships between users, components, and tasks that affect system performance

Contents

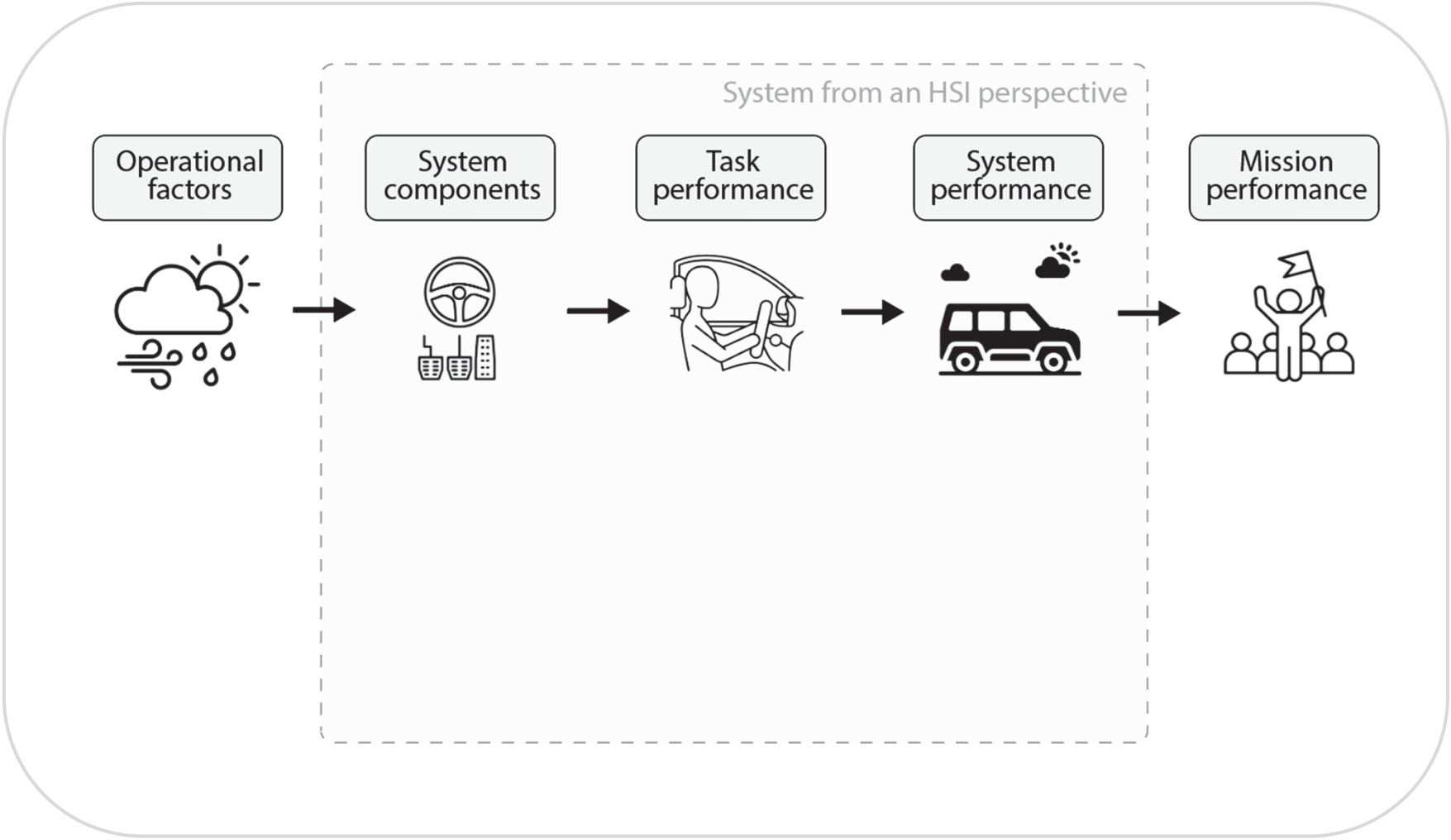
- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

HSI concepts

HSI concepts help explain how operational factors affect system performance

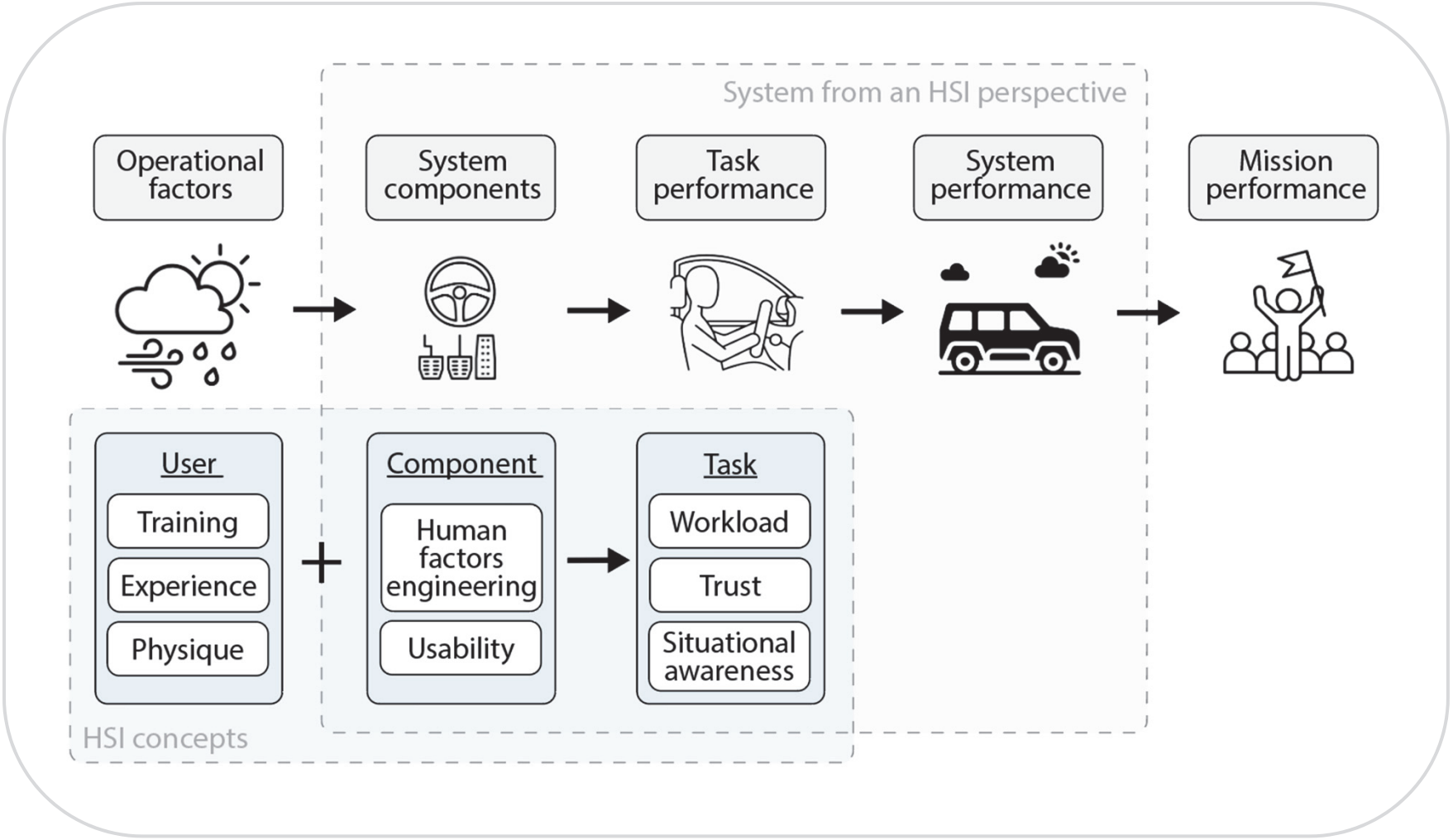
HSI – Human-Systems Interaction

HSI concepts help explain how operational factors affect system performance



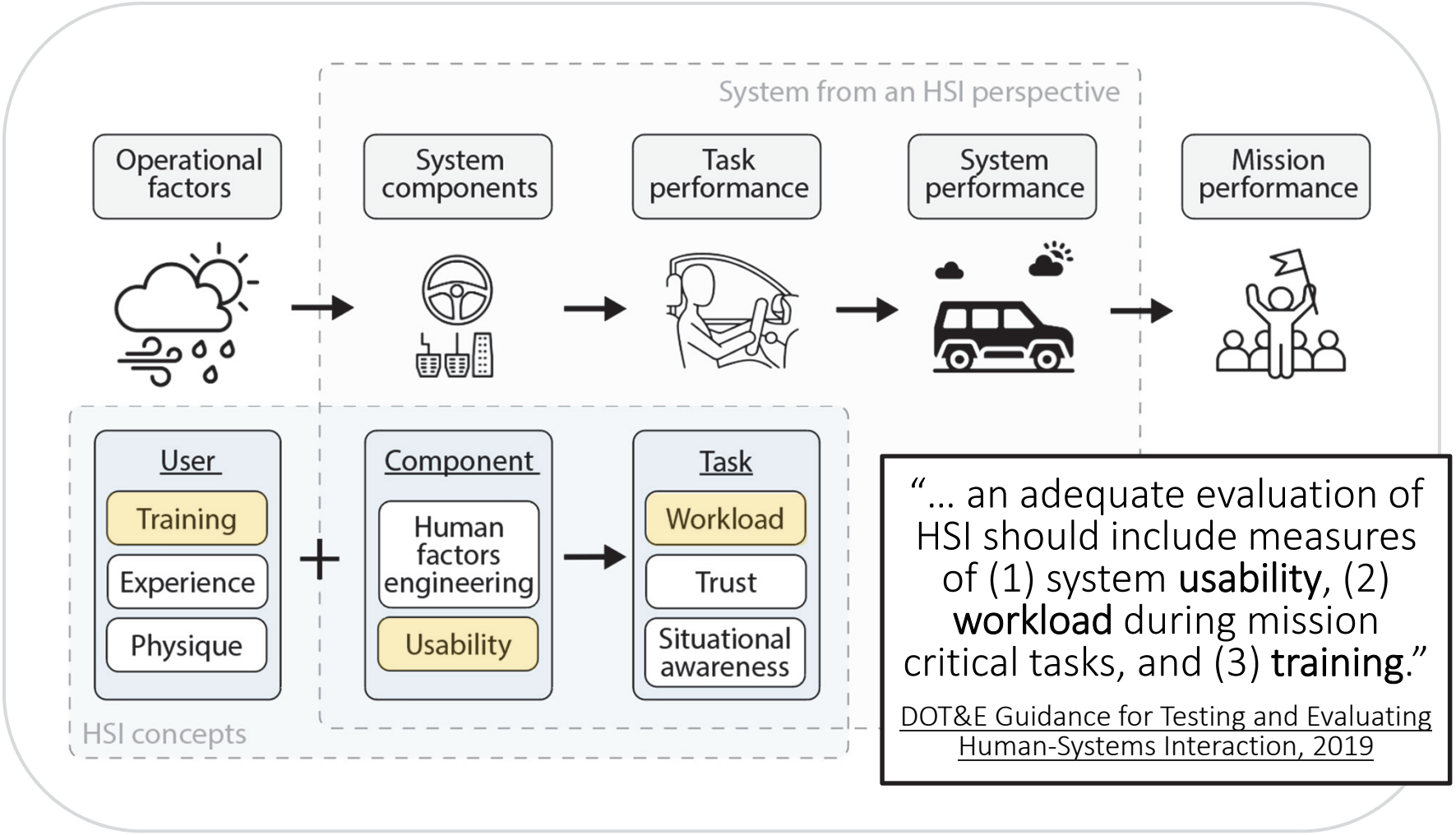
HSI – Human-Systems Interaction

HSI concepts help explain how operational factors affect system performance



HSI – Human-Systems Interaction

HSI concepts help explain how operational factors affect system performance



"… an adequate evaluation of HSI should include measures of (1) system **usability**, (2) **workload** during mission critical tasks, and (3) **training**."

DOT&E Guidance for Testing and Evaluating Human-Systems Interaction, 2019

DOT&E – Director, Operational Test & Evaluation; HSI – Human-Systems Interaction

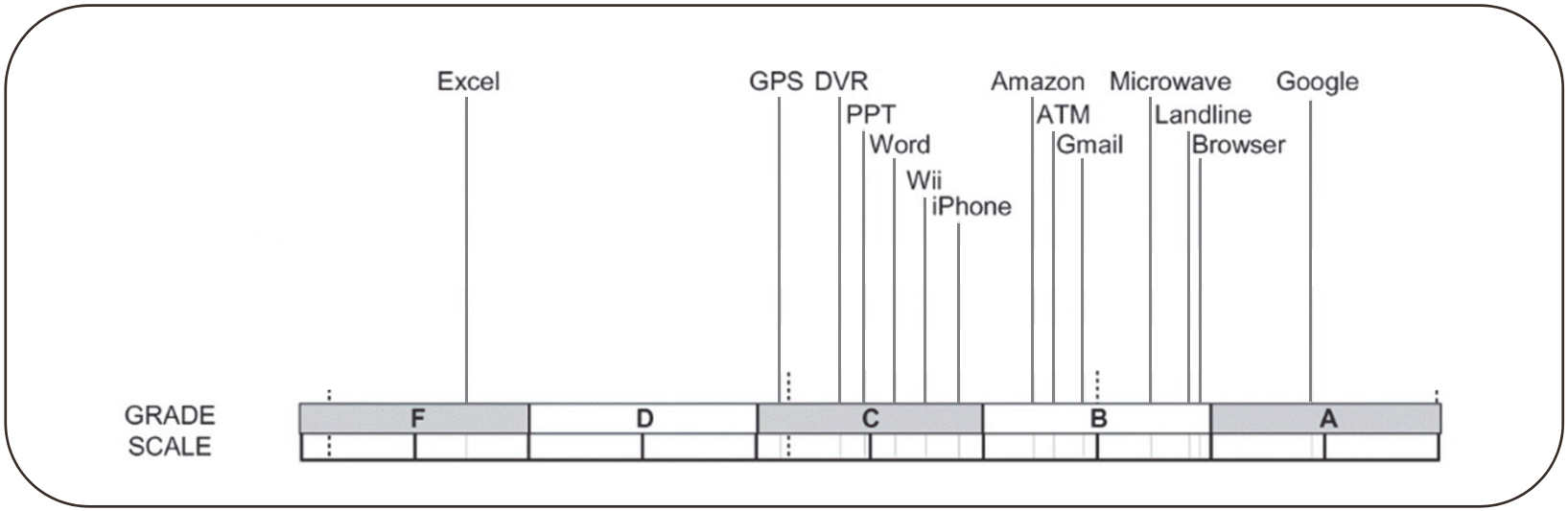
Training

- New equipment training (NET) will be ‘fielded’ with the system
  - Intended to prepare unit for operational tasks including OT
  - Typically contains a combination of lecture and hands-on training
- Evaluate whether NET prepared operators for operational tasks
  - Presentation quality is necessary but not sufficient

OT – Operational Testing

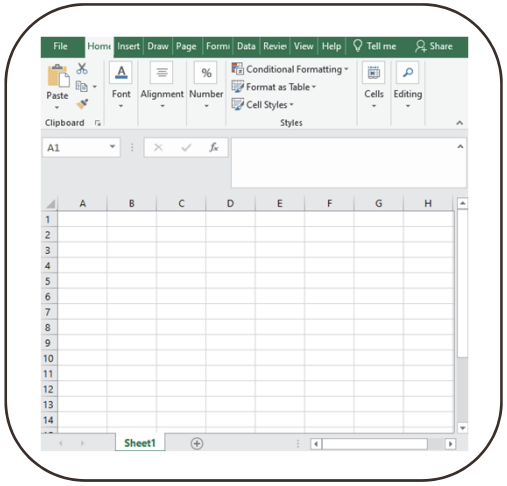
Usability

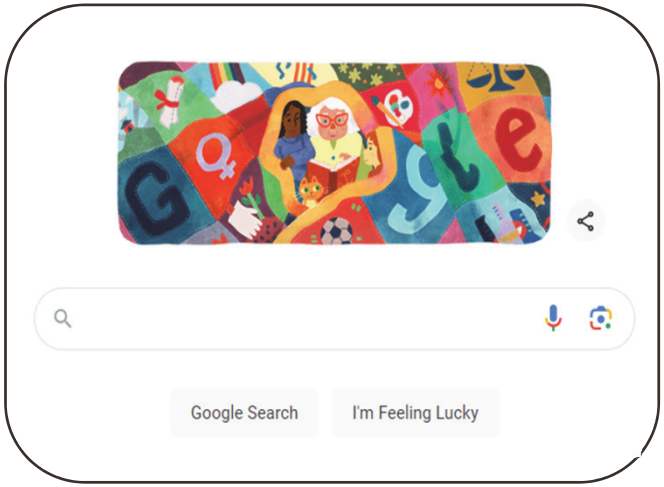
- Highly usable systems are effective, efficient, and satisfying to use. They have few design errors or inherent difficulties.
- We measure the usability of a *specific* system component.
  - Usability may vary between system components

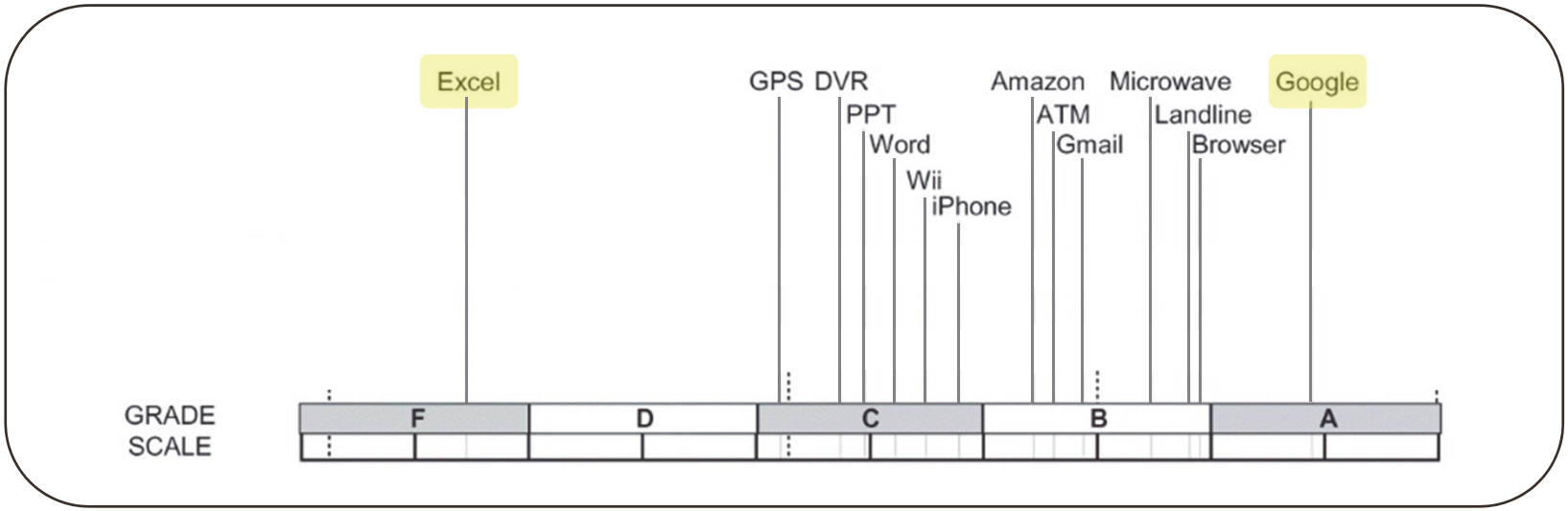


Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the System Usability Scale. International Journal of Human-Computer Interaction, 29(2), 67–76. https://doi.org/10.1080/10447318.2012.681221

Usability



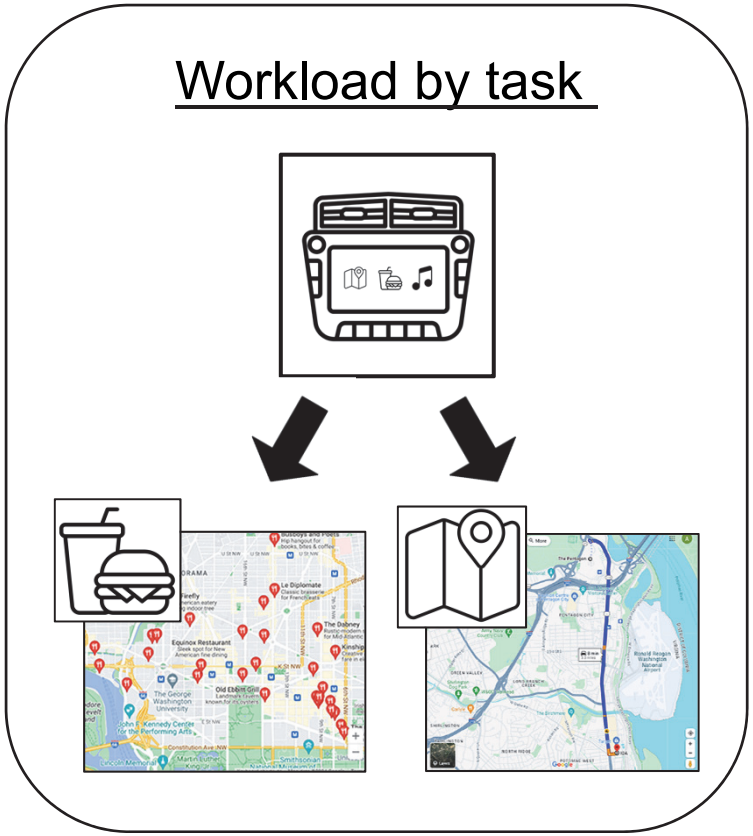


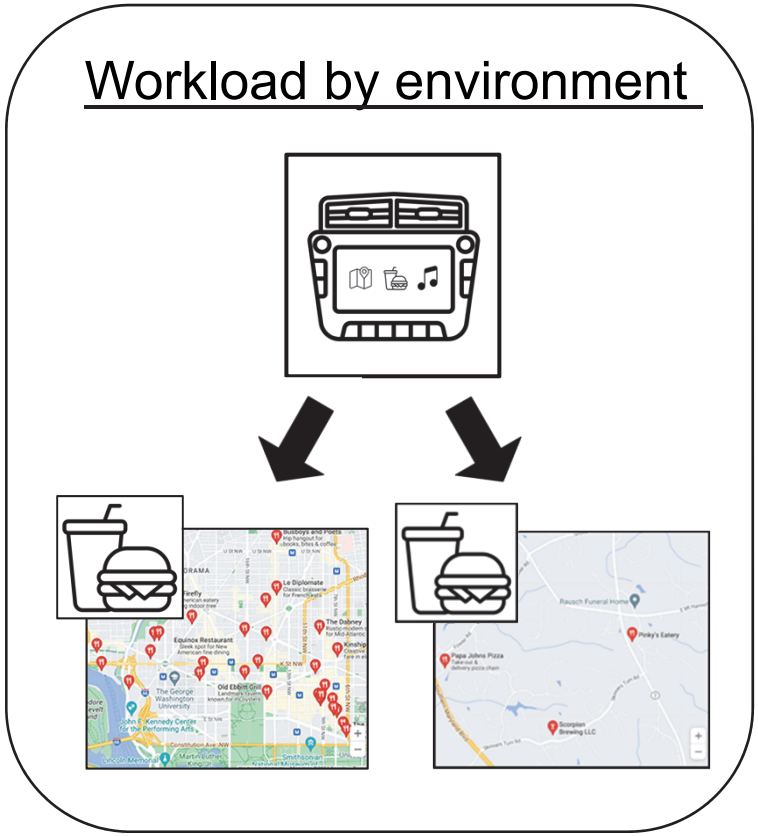


Kortum, P. T., & Bangor, A. (2013). Usability ratings for everyday products measured with the System Usability Scale. International Journal of Human-Computer Interaction, 29(2), 67–76. https://doi.org/10.1080/10447318.2012.681221

Workload

- Workload describes the effort required to perform a specific task
- Workload depends on environmental factors (e.g., stimulus density, light level, time allotment)





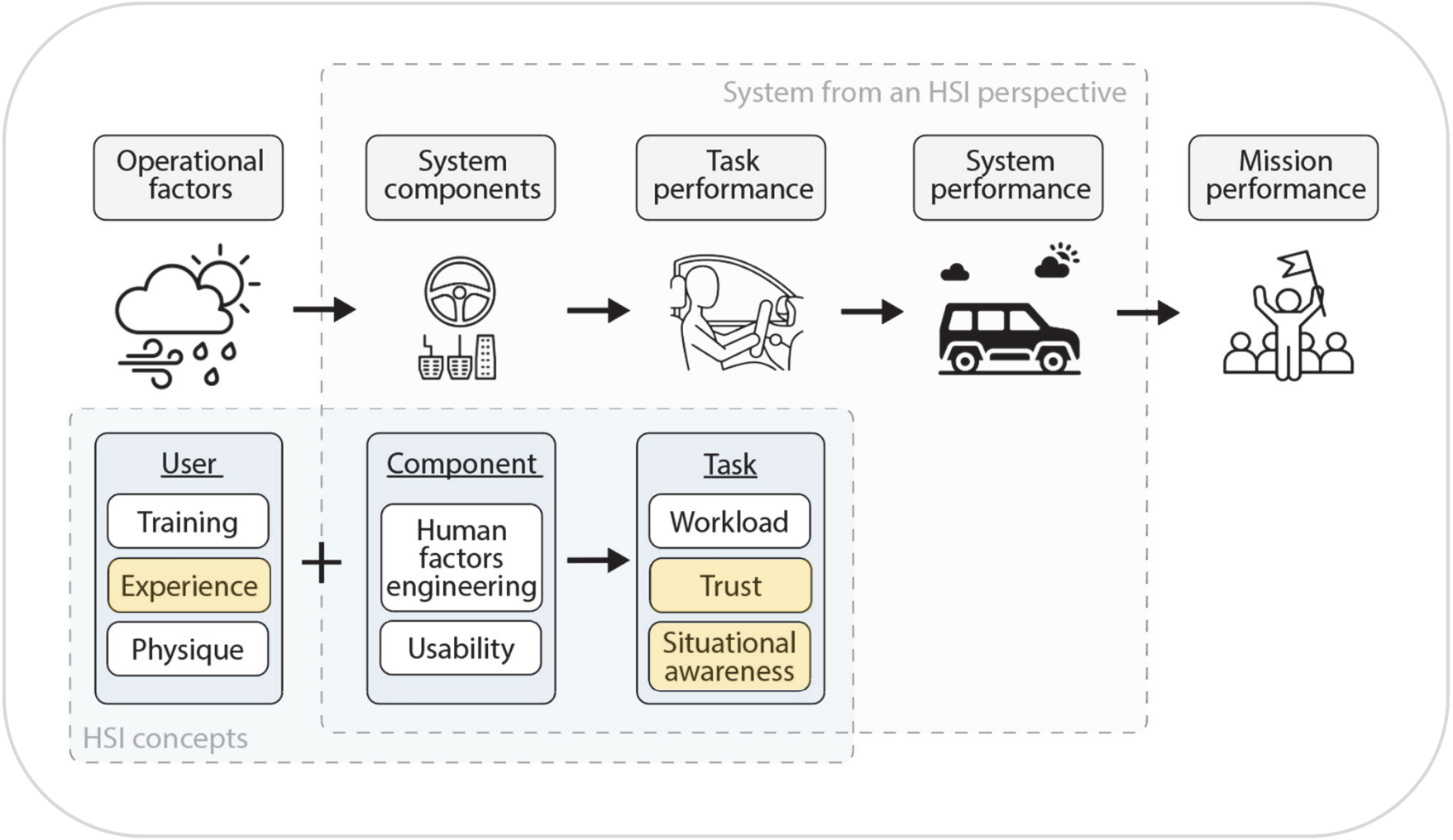
Workload

Moderate workload is best

- When workload is too low, operators become complacent
- When workload is too high, operators triage responsibilities

https://abc7.com/tesla-california-freeway-driver-asleep-caught-on-video/12776857/

HSI concepts help explain how operational factors affect system performance



HSI – Human-Systems Interaction

Experience & learning

Systems are designed for users with a specific experience profile. Accumulated experience during testing may affect performance.

Situational awareness

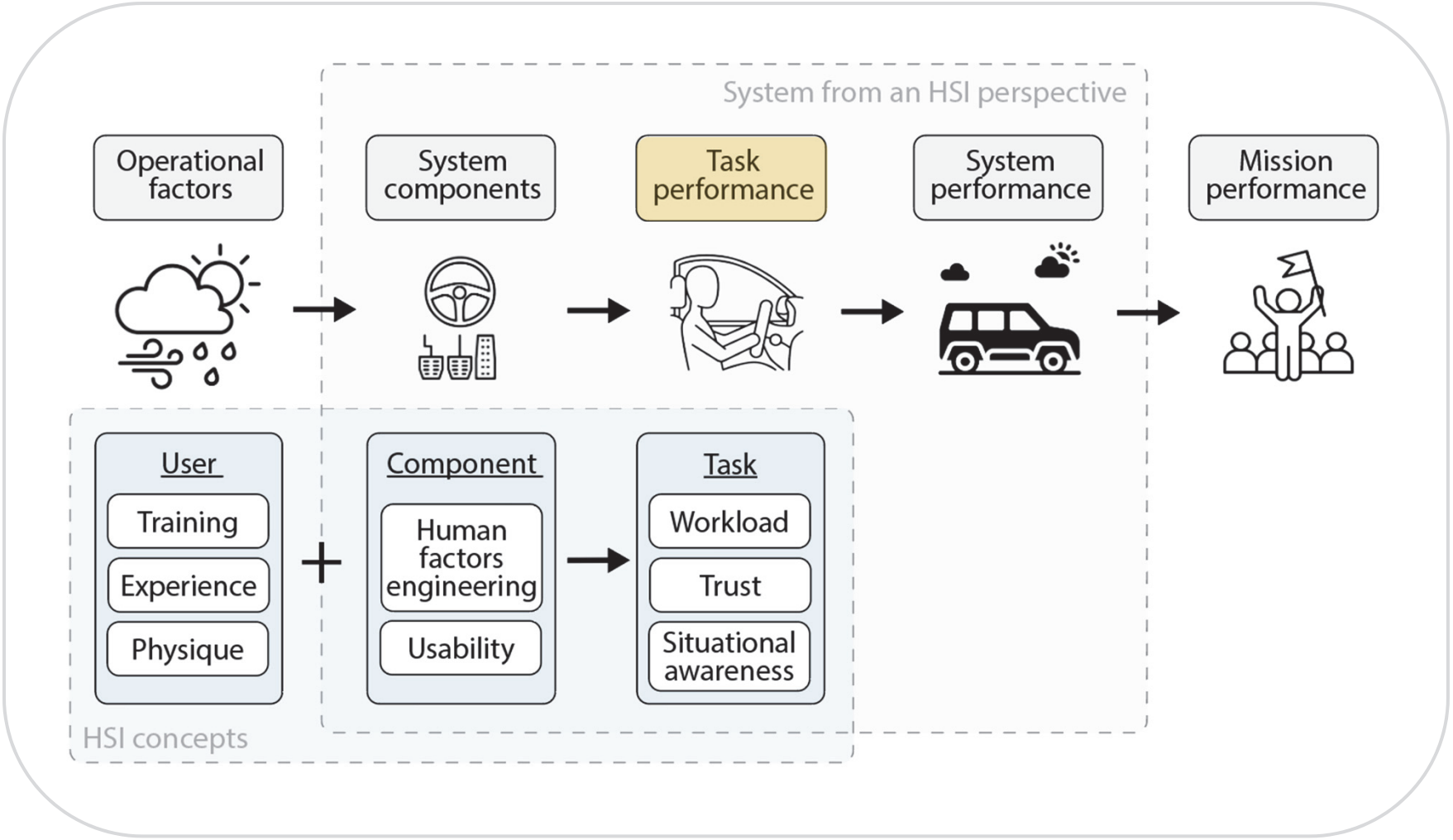
Situational awareness describes the accuracy of the user's mental model of their environment.

Trust

Trust affects whether operators will use the system.

- When trust is too high, users will overuse the system
- When trust is too low, users will underuse the system

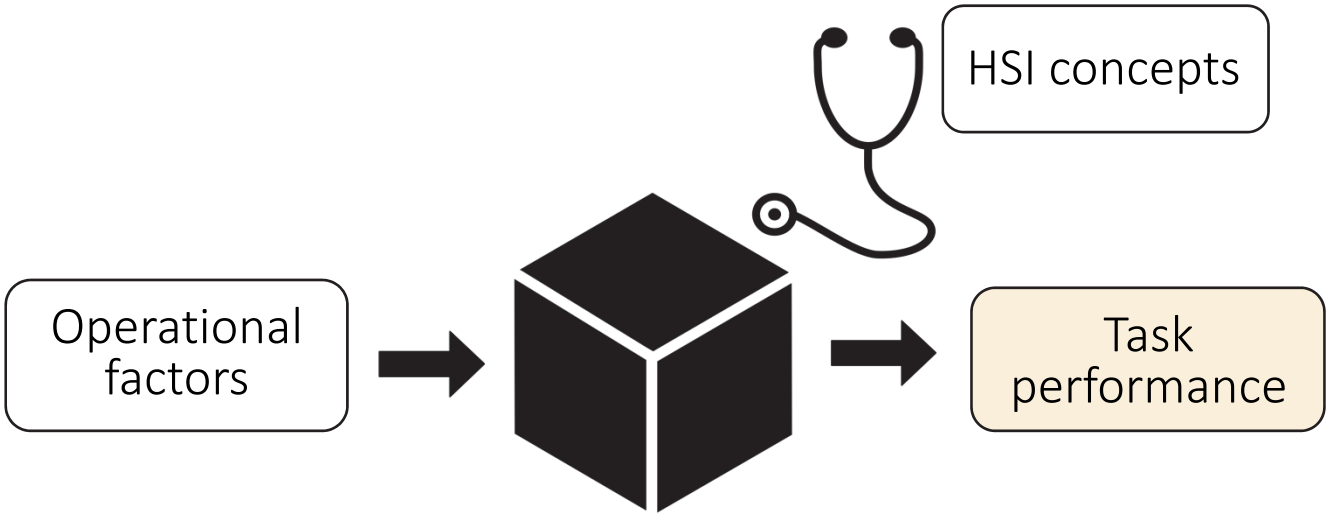
HSI concepts help explain how operational factors affect system performance



HSI – Human-Systems Interaction

Nothing in HSI makes sense without behavior

- Task performance is fundamental to operational testing
- Measuring task performance means measuring user behavior
- HSI concepts explain how operational factors affect user behavior related to task performance



HSI – Human-Systems Interaction

HSI concepts, key take-aways

- Evaluate Training, Usability, and Workload per DOT&E guidance
  - Other HSI concepts may also be important for your system
- Measure task performance behavior to understand how HSI affects system performance

DOT&E – Director, Operational Test & Evaluation; HSI – Human-Systems Interaction

Contents

- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures 
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

HSI measures

Major categories of HSI measures

Behavior

E.g., Time-to-complete, accuracy, error rate

Direct measure of operational performance

Causes of performance may be unclear

Survey

E.g., OATS, UMUX-lite, ARWES, TOAST

Provides *theoretical* context; efficient

Operational impact may be unclear

Interview

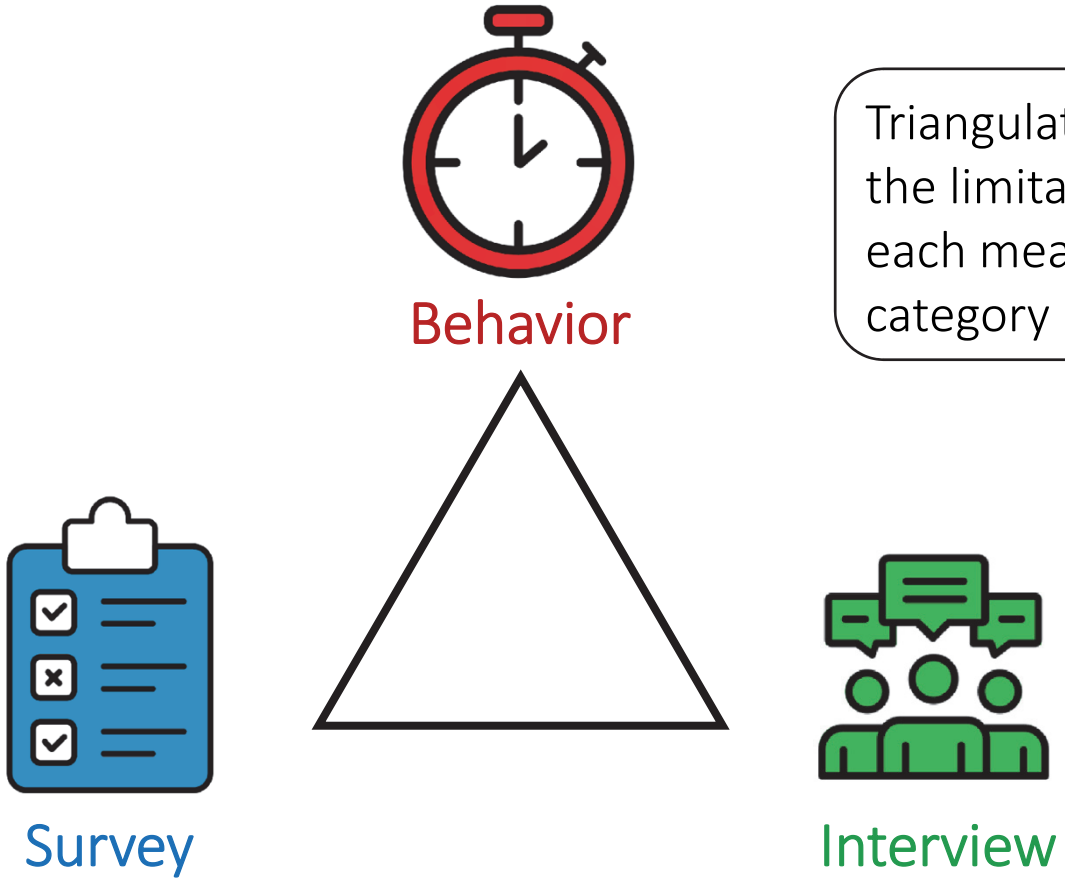
E.g., Focus groups, open-ended questions

Provides *operational* context; in-depth

Difficult to generalize to other users, situations

ARWES – Air Force Flight Test Center Revised Workload Estimate Scale; HSI – Human-Systems Interaction; OATS – Operational Assessment of Training Scale; TOAST – Trust Of Automated Systems Test; UMUX – Usability Metric for User Experience

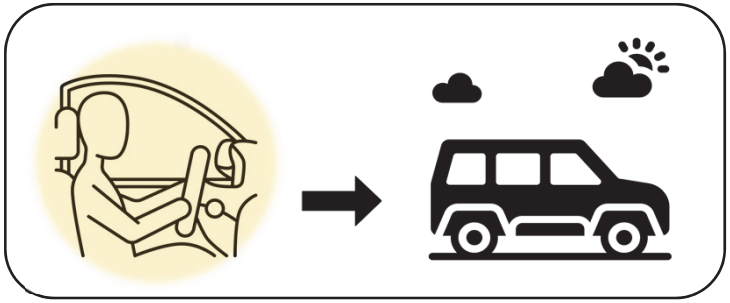
Triangulation: collect data using measures from multiple categories



Triangulation offsets the limitations of each measurement category

Behavioral measurement

- Behavioral measures provide objective, instrumented data about users' operational task performance

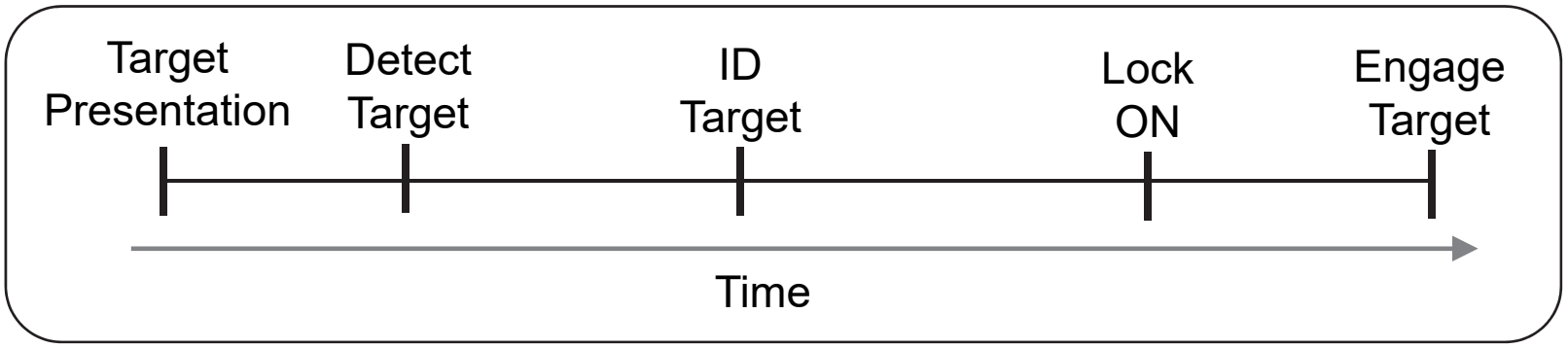


- Behavioral measurement is analogous to instrumented measurement of physical systems, with a similar planning process:
  1. Define the research question in terms of user behavior
  2. Define the specific behavioral variable of interest
  3. Identify appropriate instruments for measuring the behavioral variable
  4. Incorporate the appropriate instruments into the operational test
  5. Collect and process the behavioral data

Behavioral measurement example

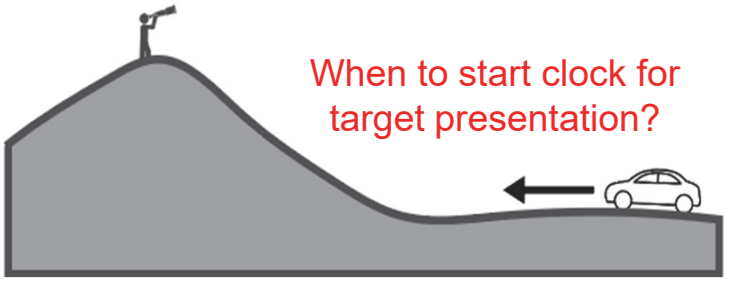
Usability of Javelin Lightweight Command Launch Unit (LWCLU)

- Does LWCLU improve timeliness of engagement process over legacy system?
- Instrumentation
  - Device logs
  - Verbal report & stopwatch
- Generate a timeline for how quickly a task was completed



Behavioral measurement challenges

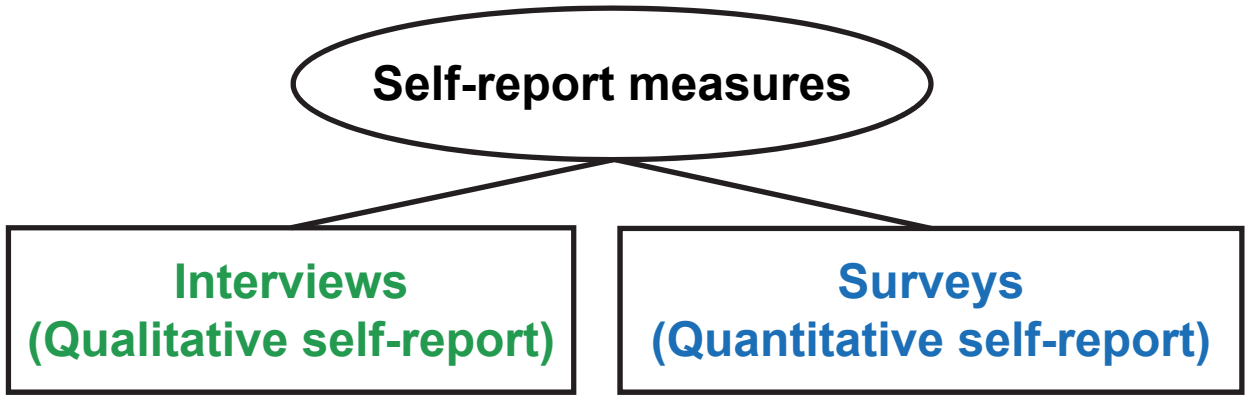
- Designing operational tasks to support data analysis
- Defining observable behaviors related to system performance
- Instrumenting users during an operational test
  - Automation is preferred
  - Minimize disruptions to operational realism



How to measure user thoughts (e.g., when user detects target)?

Self-report measures

- Self-report measures provide access to user thoughts
- Well suited for questions about opinion
  - E.g., which device do you prefer?
- Less appropriate for questions about behavior
  - E.g., how mobile were you?
- Poorly suited for questions about observable facts
  - E.g., is Device A heavier than Device B?

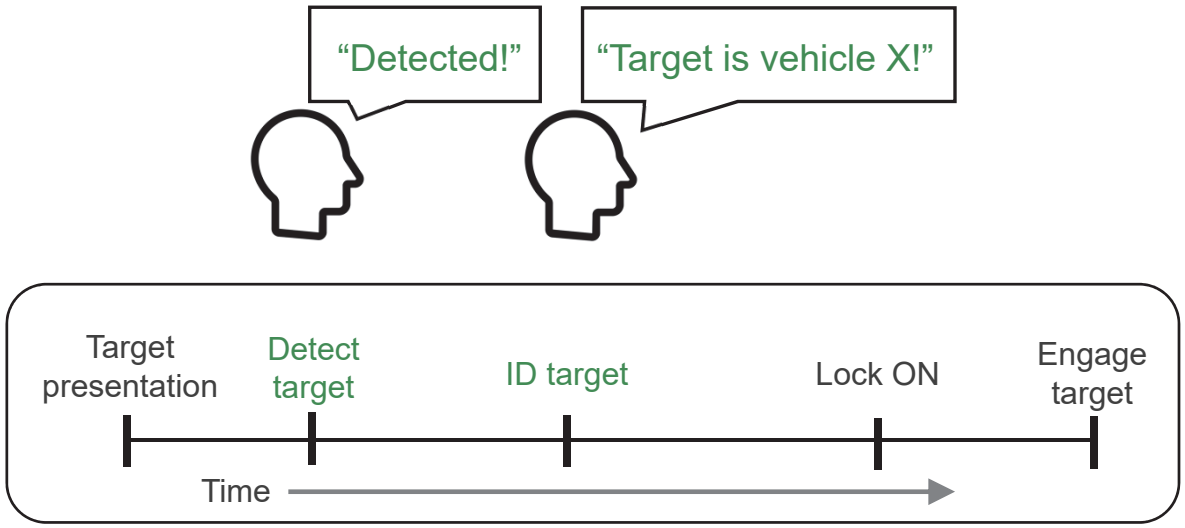


Interviews

- In-depth feedback via open-ended questions
  - **Focus groups**: 5–10 users engaging with one another and a moderator
  - **Interviews**: 1 user engaging with an interviewer
  - **Open-ended survey questions**: users responding to a survey
- Provide invaluable access to user insights
  - Test unit may be the only operators with this experience
  - The system under test was specifically designed for similar operators
- Open-ended format is ideal for problem discovery

Interview example #1

Usability of the Javelin Light Weight Command Launch Unit (LWCLU)



Interview example #2

Usability of the Javelin Light Weight Command Launch Unit (LWCLU)

In a focus group:

Moderator: "A couple times while you were out there you would hand [the LWCLU] off to someone who hadn't used it and there was a quick over-the-shoulder [instruction]. Could you talk me through it and if it was effective?"

Test player: "He handed me the LWCLU… He gave me a quick run down of the buttons and what they do. I think it's pretty straight forward. I like the layout of the system as well."

Moderator: "How long would you say it took?"

Test player: "I would say within probably [specific time] he was able to put me on target."

Interview challenges

- Interviews are sensitive to moderator skill
  - Topics and some questions should be planned in advance
  - Must promote an environment that is focused and participatory
- Interviews can be resource intensive for testers and analysts
- Open-ended survey questions are a compromise
  - They collect qualitative data but cannot be immediately followed up
- Must be collected soon after the event

Surveys

- More statistical but less exhaustive than interviews
- Structured format lends itself to rigorous validation
- May come with benchmarks that contextualize absolute scores
- Can easily be administered to many individuals

Survey administration example

Usability of the Light Weight Command Launch Unit (LWCLU)

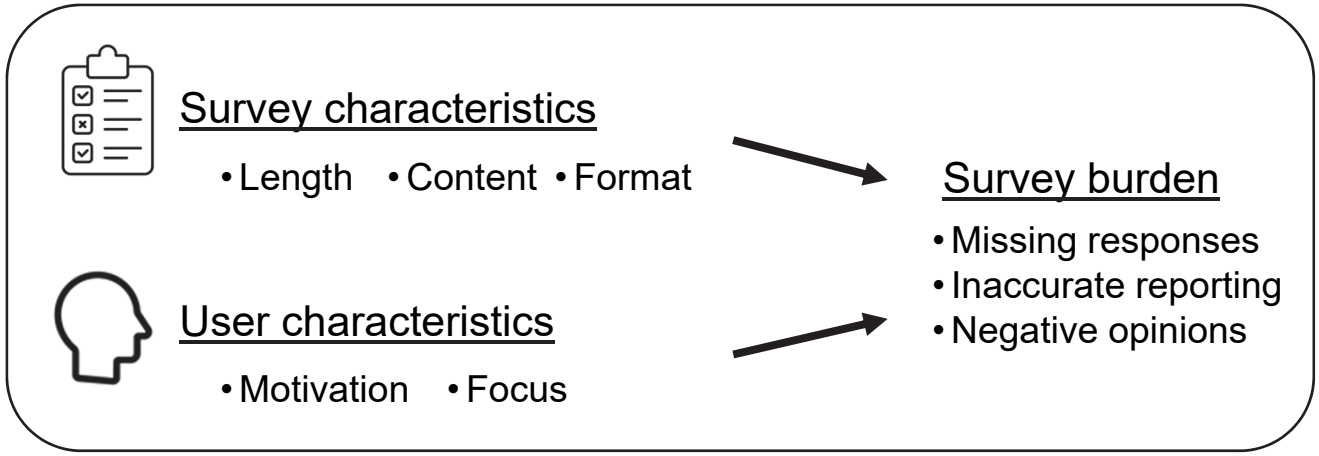
- Surveys were used to capture scores describing subjective usability of the LWCLU under operational conditions
- A short survey was administered after missions according to design of experiments
  - Usability Metric for User Experience LITE (UMUX-LITE)

Item		Strongly Disagree						Strongly Agree
1	[This system's] capabilities meet my requirements.	1	2	3	4	5	6	7
2	[This system] is easy to use.	1	2	3	4	5	6	7

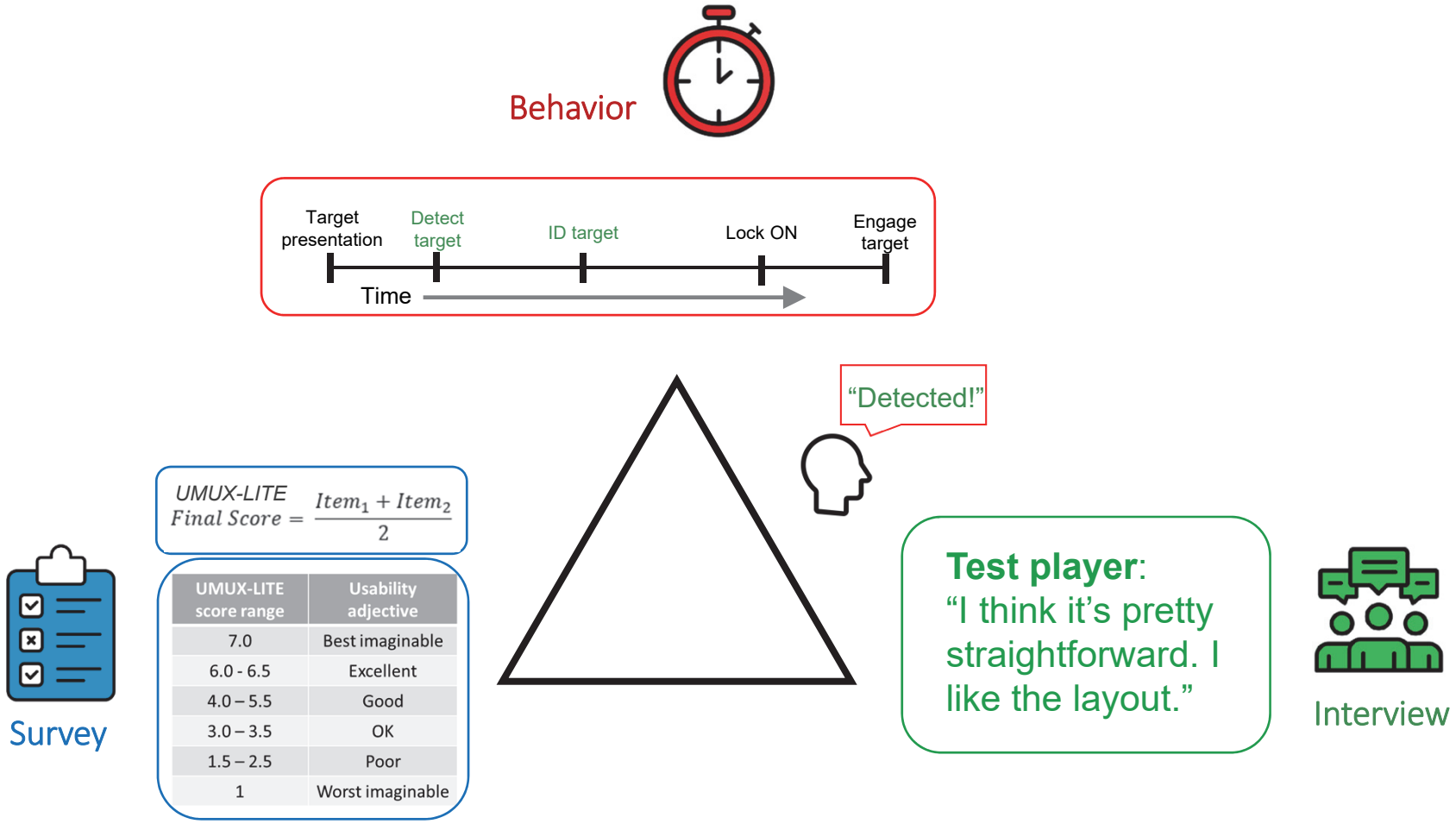
$$Final\ Score = \frac{Item_1 + Item_2}{2}$$

Survey challenges

- Survey validation process requires time, effort, money
  - User responses are sensitive to how questions are worded
  - Overall scores depend on question selection
- Results are sensitive to administration process
  - Must be administered soon after the event
  - Environment should be quiet, focused, motivating
- User fatigue can add noise to survey data:



Triangulating the usability of the LWCLU



LWCLU – Lightweight Command Launch Unit

HSI measures, key take-aways

- Any HSI concept can be measured multiple ways
- Use triangulation to strengthen your analysis
  - Behavioral data, interview data, survey data
- Use rigorous methodology to reduce measurement error
  - Automated measures, trained moderators, validated surveys
  - Collect data during or immediately after event
- Understand the provenance of your data
  - All memories are made up, some are true
  - Self-reported facts are *beliefs*

Contents

- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

HSI measures: validated surveys

"Validated surveys reduce error and should be used in lieu of custom-written questions whenever possible. A repository of [validated surveys] is available on the DOT&E website."

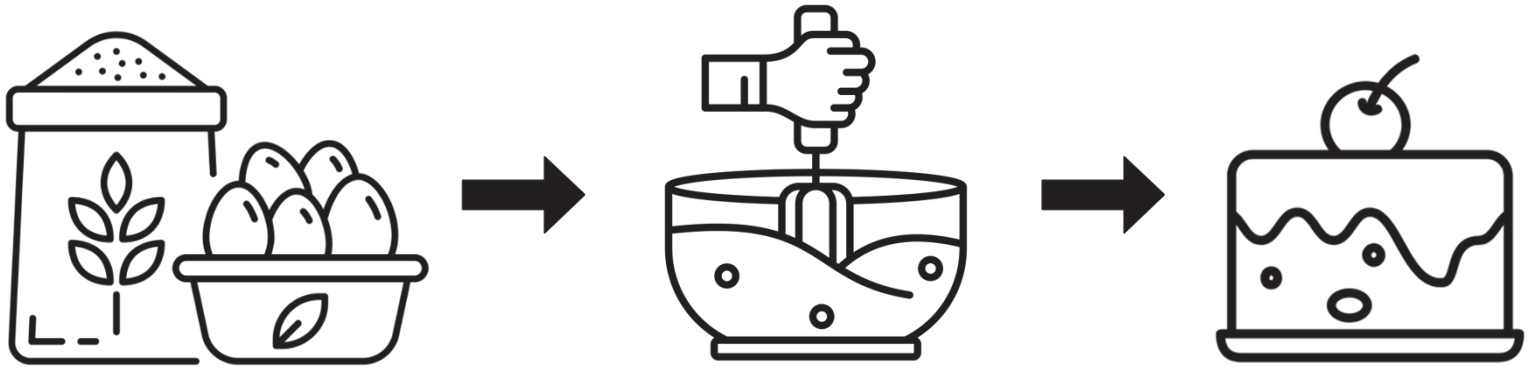
Guidance for Testing and Evaluating Human-System Interaction
2019 September, DOT&E memo

See also: testscience.org/validated-scales-repository-intro

Validated surveys are like recipes

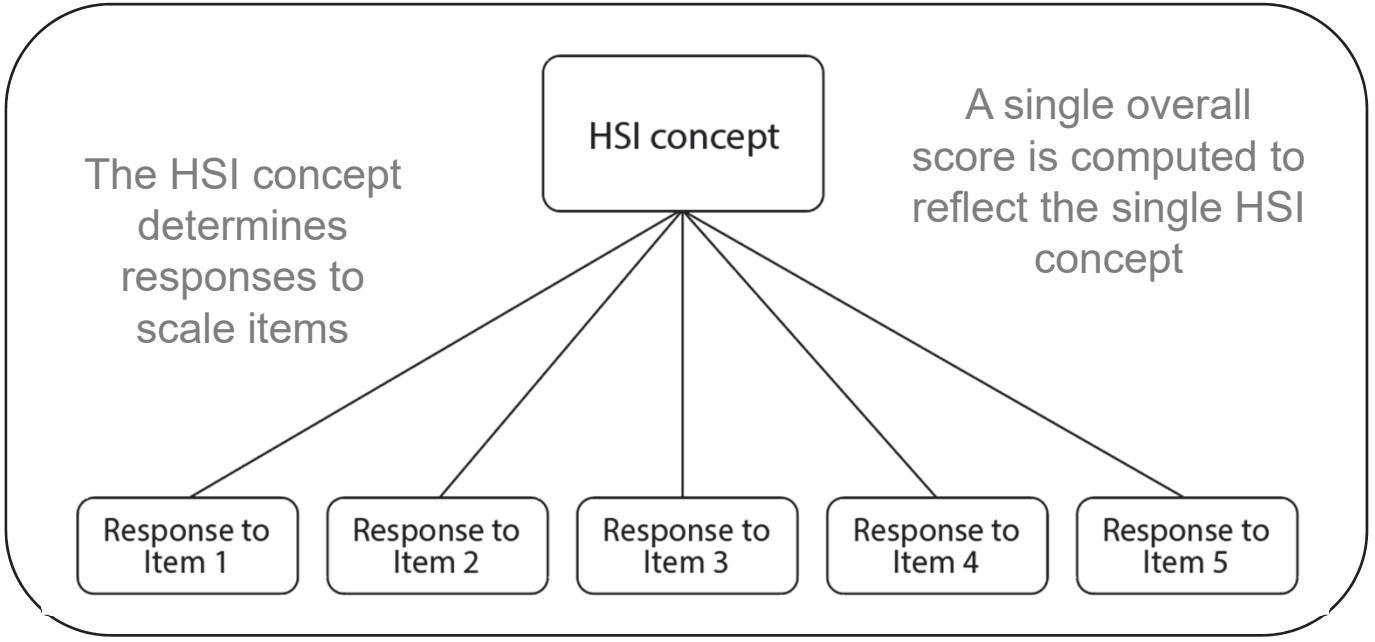
Ingredients must be tested for quality

Combination must be tested to ensure it reliably produces a useful product



Validated survey scores quantify HSI concepts as latent constructs

- Modern surveys are designed to measure latent constructs
  - A latent construct (e.g., intelligence) can only be inferred indirectly
  - Questions are selected for their relationship with the construct
  - Surveys are scored according to the information provided by each question



HSI – Human-Systems Interaction

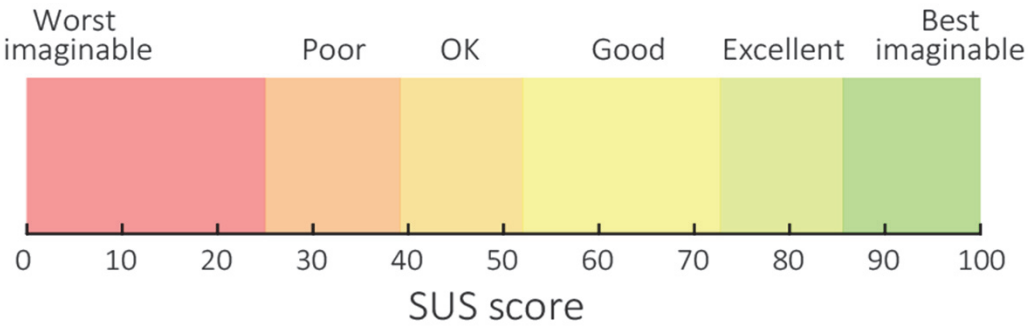
Validated surveys have been tested

Individual items are tested to ensure clarity, reliability, accuracy

SUS item 3	Strongly Disagree				Strongly Agree
I thought the system was easy to use.	1	2	3	4	5

Overall scales are tested to ensure relevance and meaningfulness

$$FinalScore = \big((Item_1 - 1) + (Item_2 - 1) + (Item_3 - 1) + (Item_4 - 1) + (Item_5 - 1) + (Item_6 - 1) + (Item_7 - 1) + (Item_8 - 1) + (Item_9 - 1) + (Item_{10} - 1)\big) * 2.5$$



SUS – System Usability Scale

Some validated surveys approved by DOT&E

Training effectiveness	Number of questions
Operational Assessment of Training Scale (OATS)	6
Training Structure Assessment (TSA, aka DSoT)	8

Usability	Number of questions
System Usability Scale (SUS)	10
Usability Metric for User Experience LITE (UMUX-LITE)	2

Some validated surveys approved by DOT&E

Workload	Number of questions
NASA Task Load Index (NASA TLX)	6–21
AFFTC Revised Workload Est. Scales (ARWES) / Crew Status Survey (CSS)	1

User Trust in Systems	Number of questions
Trust Of Automated Systems Test (TOAST)	9

Validated surveys, key take-aways

- Validated surveys have been tested to ensure that they produce reliable and accurate results
- DOT&E has approved commonly used validated scales
  - See: testscience.org/validated-scales-repository-intro
- Other validated scales may exist for your research question
  - Reach out to the HSI team

Contents

- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

HSI measures: custom surveys

"However, validated measures do not yet exist for some HSI concepts and custom-written questions may be required. If testers plan to use custom questions, they must be pre-tested with the target audience."

Guidance for Testing and Evaluating Human-System Interaction
2019 September, DOT&E memo

A custom survey is like a list of ingredients

Ingredients must be evaluated for quality

Combination must be evaluated to ensure it reliably produces a useful product

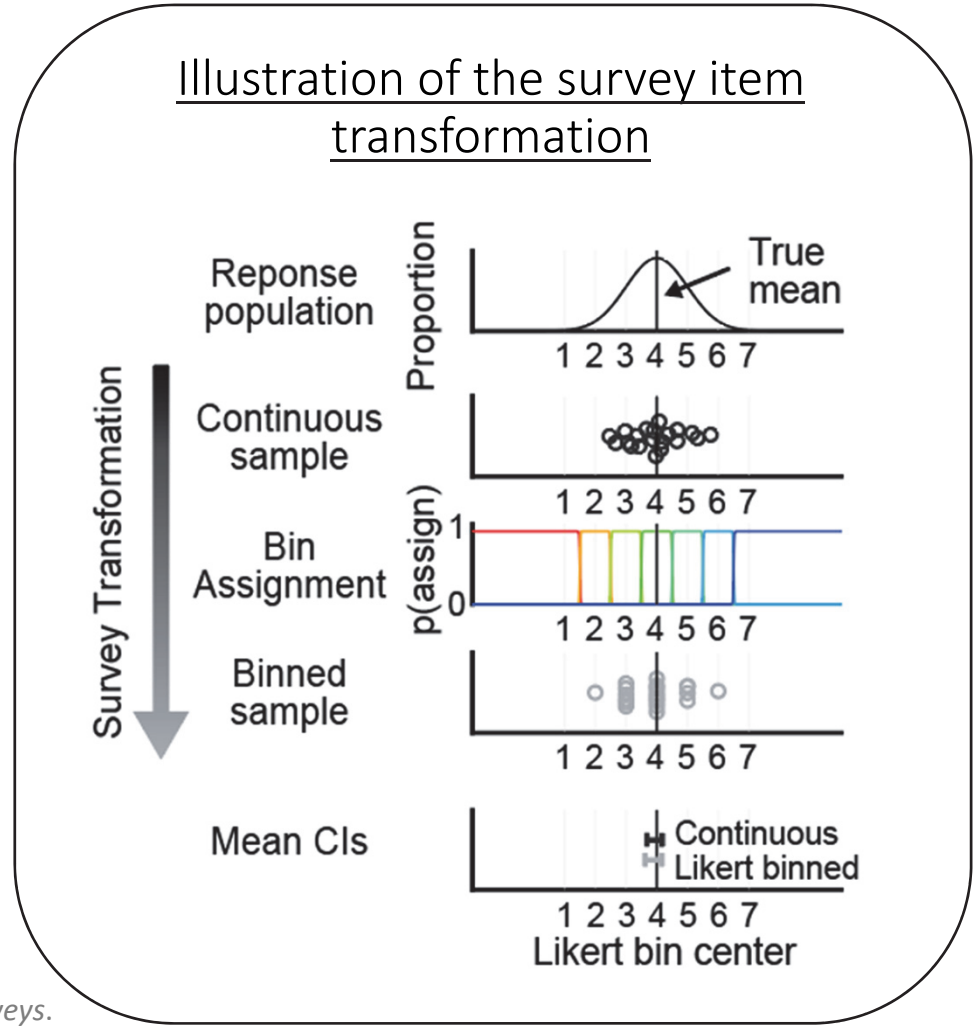
→ ? → ?

Custom surveys have not been tested

- They rarely attempt to measure HSI concepts
- They are often poorly designed
  - Not tested to ensure reliability or accuracy
  - May misrepresent user opinions and invalidate statistical analyses
- Testers should consider replacing or improving custom surveys
  - Replace: use validated surveys or qualitative methods (e.g., interviews)
  - Improve: edit individual items and then encourage pretesting (i.e., red teaming)

Surveys transform user opinions into numerical responses

- Likert survey items assign numbers to user opinions
- Poorly designed surveys can introduce unintended changes to the data
- These changes cannot be undone by adding more subjects or by using alternative statistics

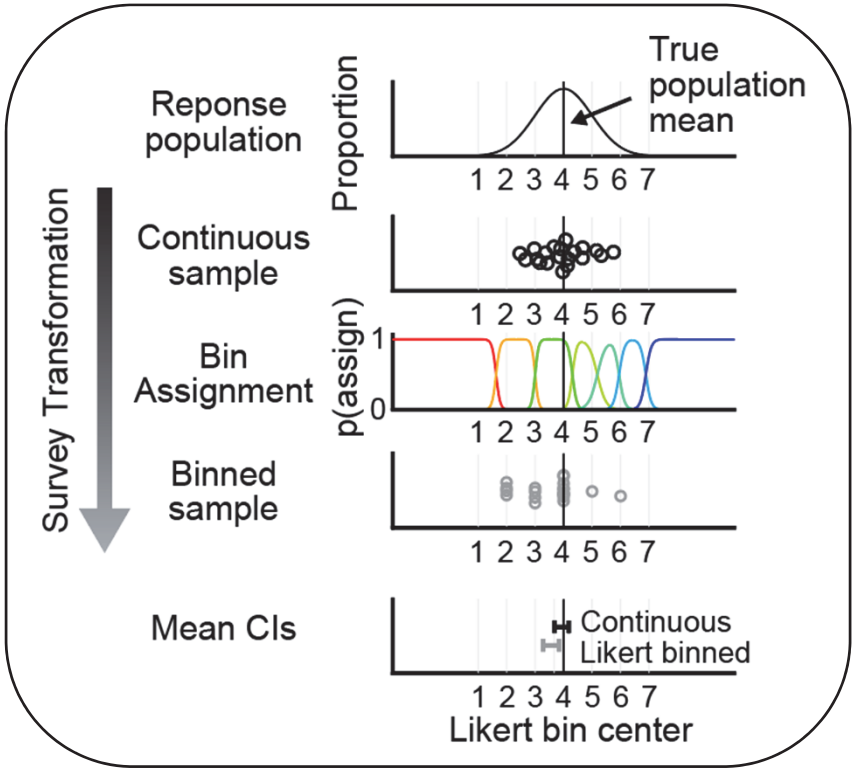


Miller, Adam M. 2024. *DATAWorks 2024: Statistical Advantages of Validated Surveys over Custom Surveys*.

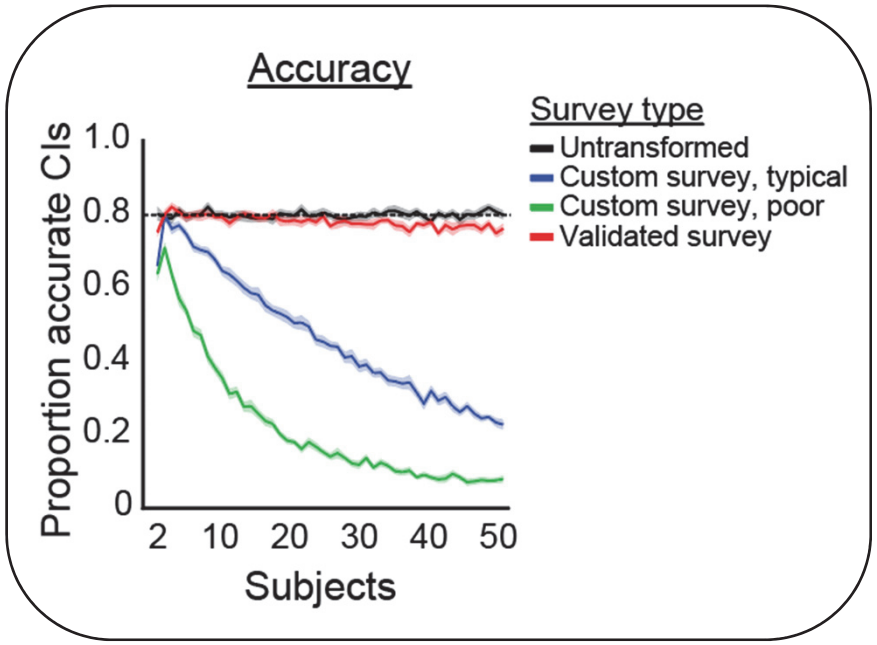
CI – Confidence Interval

Poorly designed surveys distort user data

Poor survey design can introduce unintended changes to user data



Poor survey design invalidates confidence intervals

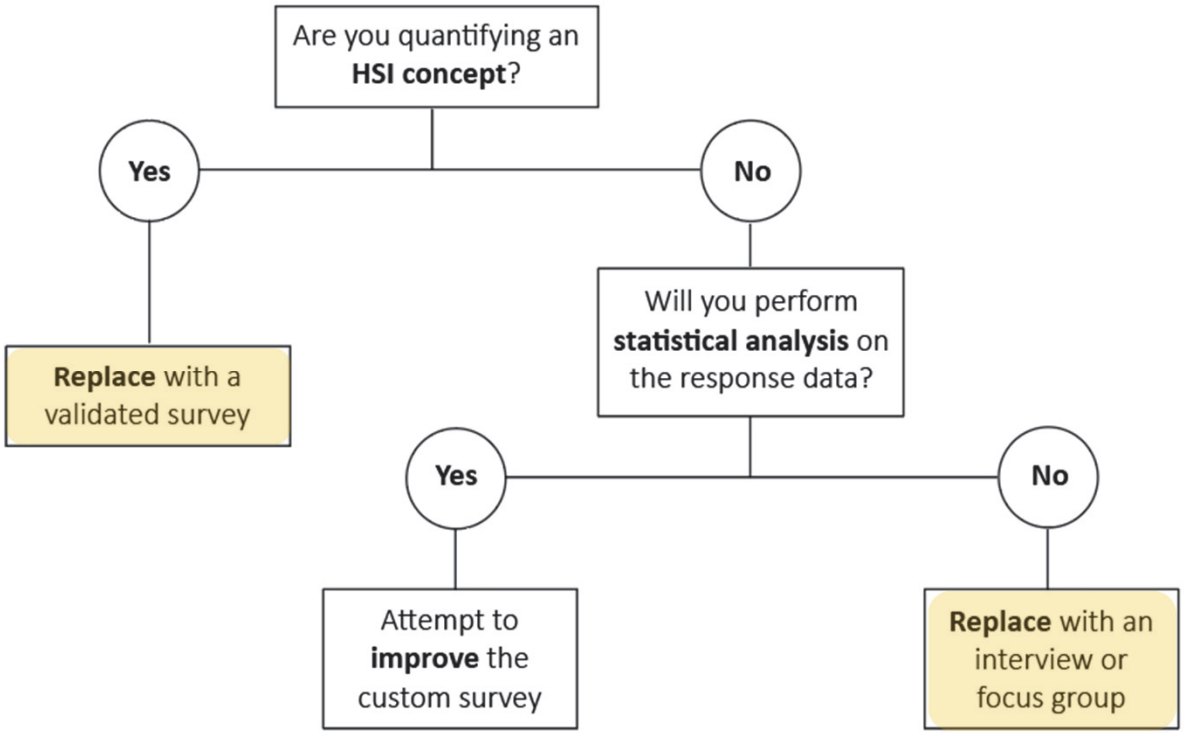


Miller, Adam M. 2024. *DATAWorks 2024: Statistical Advantages of Validated Surveys over Custom Surveys*.

CI – Confidence Interval

Option 1: Replace custom surveys

- Only validated surveys should be used to measure HSI concepts
- An interview is a more rigorous way to "just ask questions"



HSI – Human-Systems Interaction

Option 2: Improve custom surveys: item questions

Common problems with item questions

- Asking multiple questions in one
  Avoid: "The screen was bright and easy to touch."
- Using confusing language
  Avoid: "The agency transfer controller switched tasks fluidly."
- Leading the witness
  Avoid: "This was the best system ever."
- Asking an impossible question
  Avoid: "I detected all targets."
- Measuring an observable fact
  Avoid: "Did the computer crash?"

Potential solutions

- Ask two questions
  Try: "The screen was bright."
  "The screen was easy to touch."
- Use familiar language
  Try: "It was easy to switch between tasks."
- Use clear language
  Try: "I prefer this system to my current solution."
- Ask about user experience
  Try: "I was rarely surprised by targets."
- Ask about user experience
  Try: "I was able to complete my tasks without interruption."

Option 2: Improve custom surveys: item responses

Common problems with item responses	Potential solutions
• Using non-Likert response categories	Try: 7pt Likert scale
• Failing to provide a neutral response	Try: 'Neutral' response
• Failing to allow users to opt-out	Try: 'Not applicable' response

Custom surveys, key take-aways

- Statistical estimates of sample population parameters may be inaccurate when computed from custom survey data
- If statistics are not needed, consider using interviews instead
- If you must use custom surveys, be sure to follow survey design best practices

Contents

- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting
- Reviewing test plans for HSI
- Resources

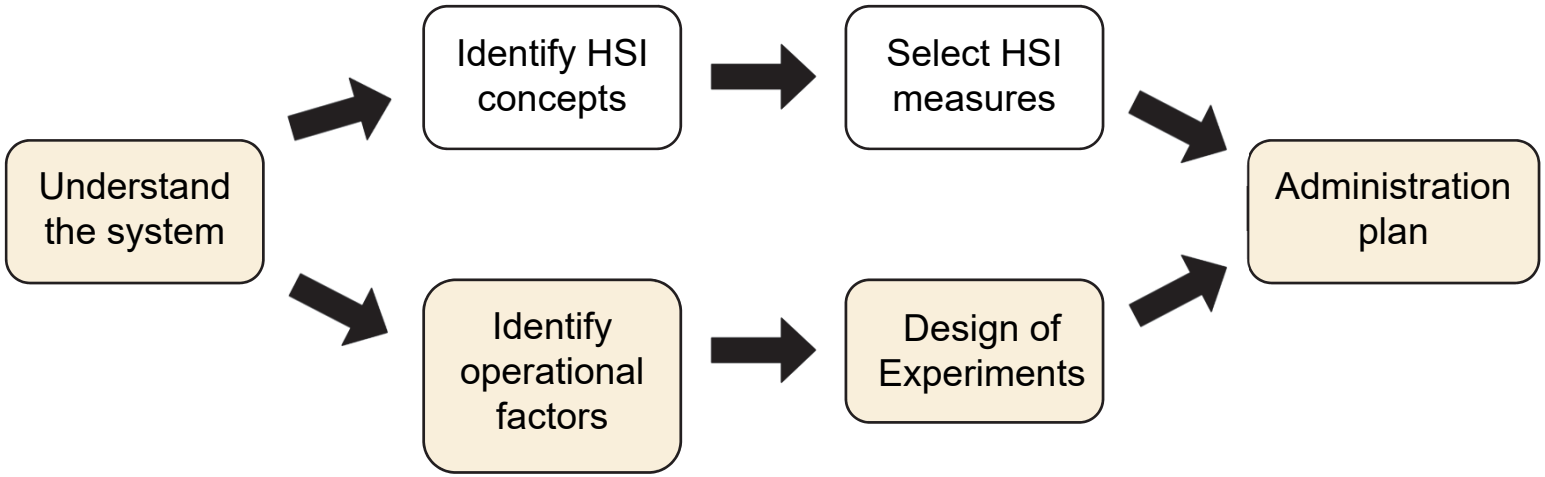
HSI design of experiments

HSI design of experiments

Design of Experiments (DOE)

Measure HSI concepts over the operational envelope

- Identify factors that affect system performance
- Design a test with the analysis you want in mind
- Determine how much data you need
- Administer HSI measures across relevant test conditions

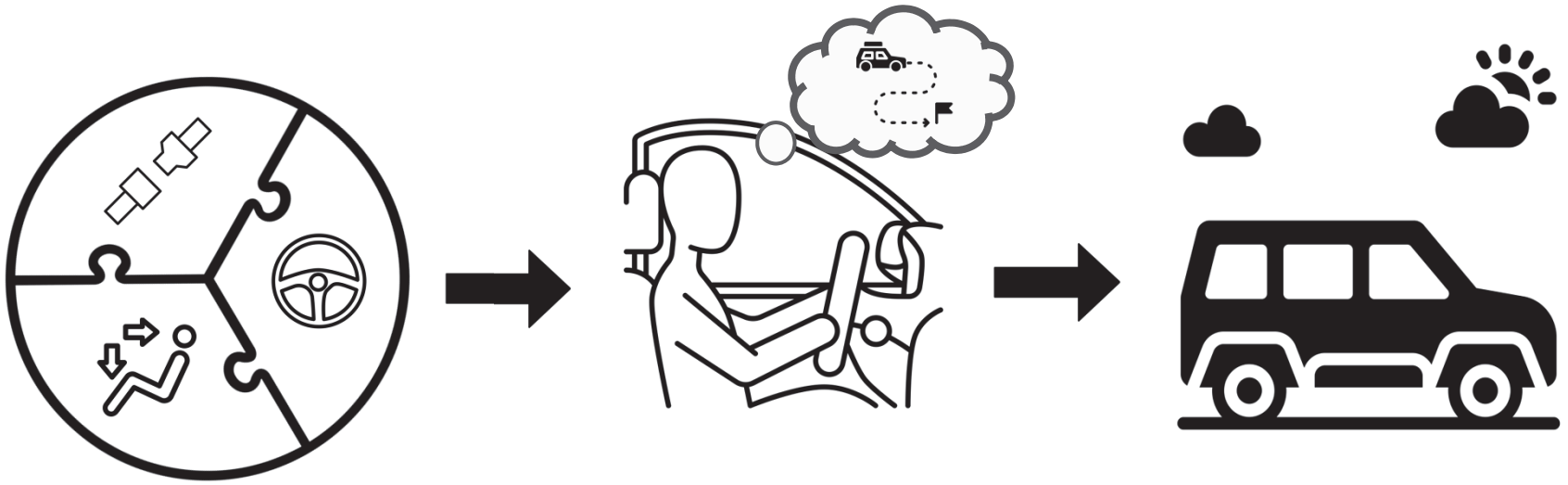


Review: understand the system from an HSI perspective

Systems are made up of components

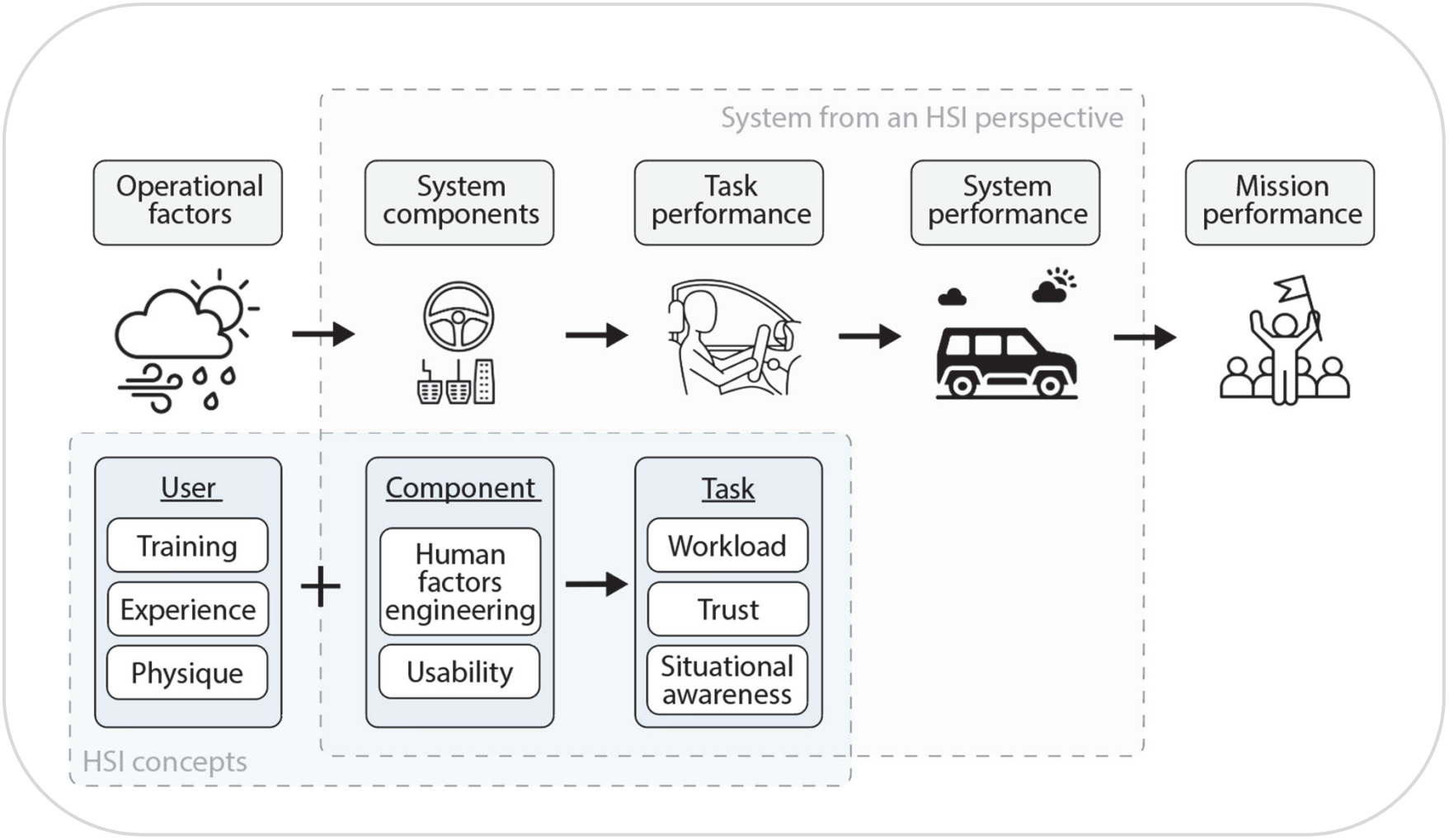
Users perform tasks with components

Tasks support system performance



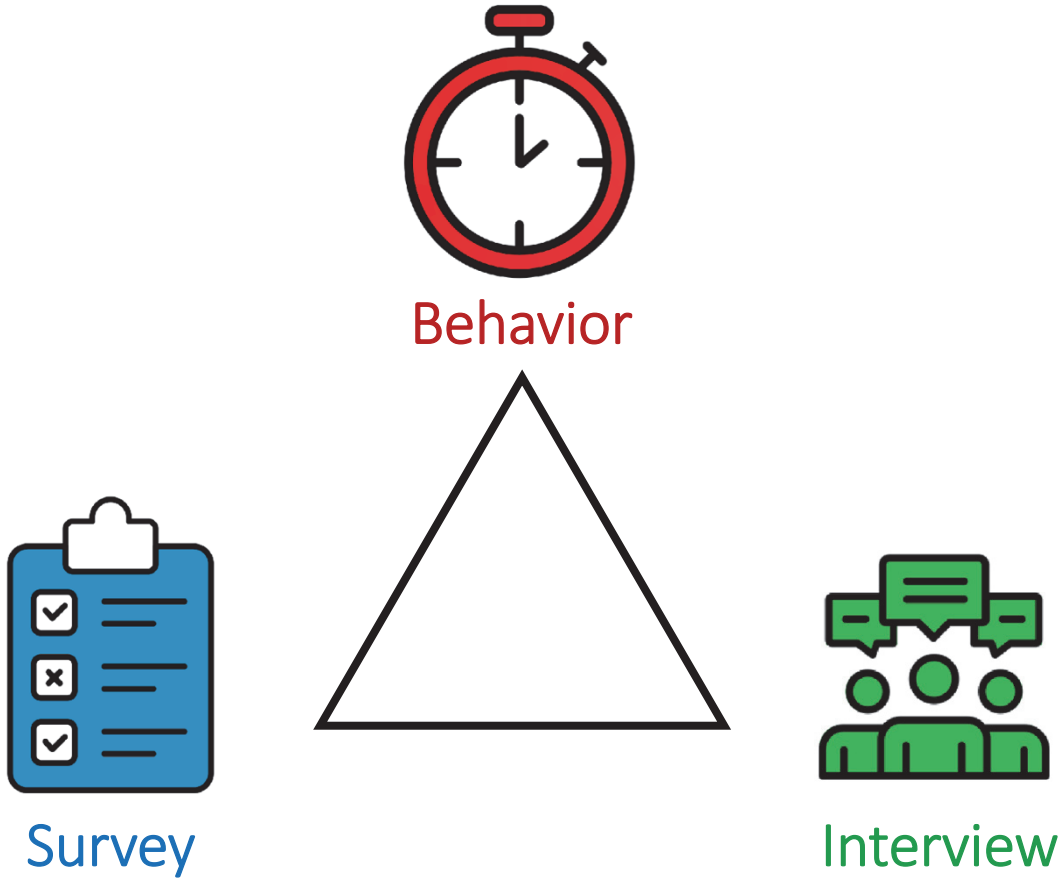
HSI – Human-Systems Interaction

Review: identify relevant HSI concepts



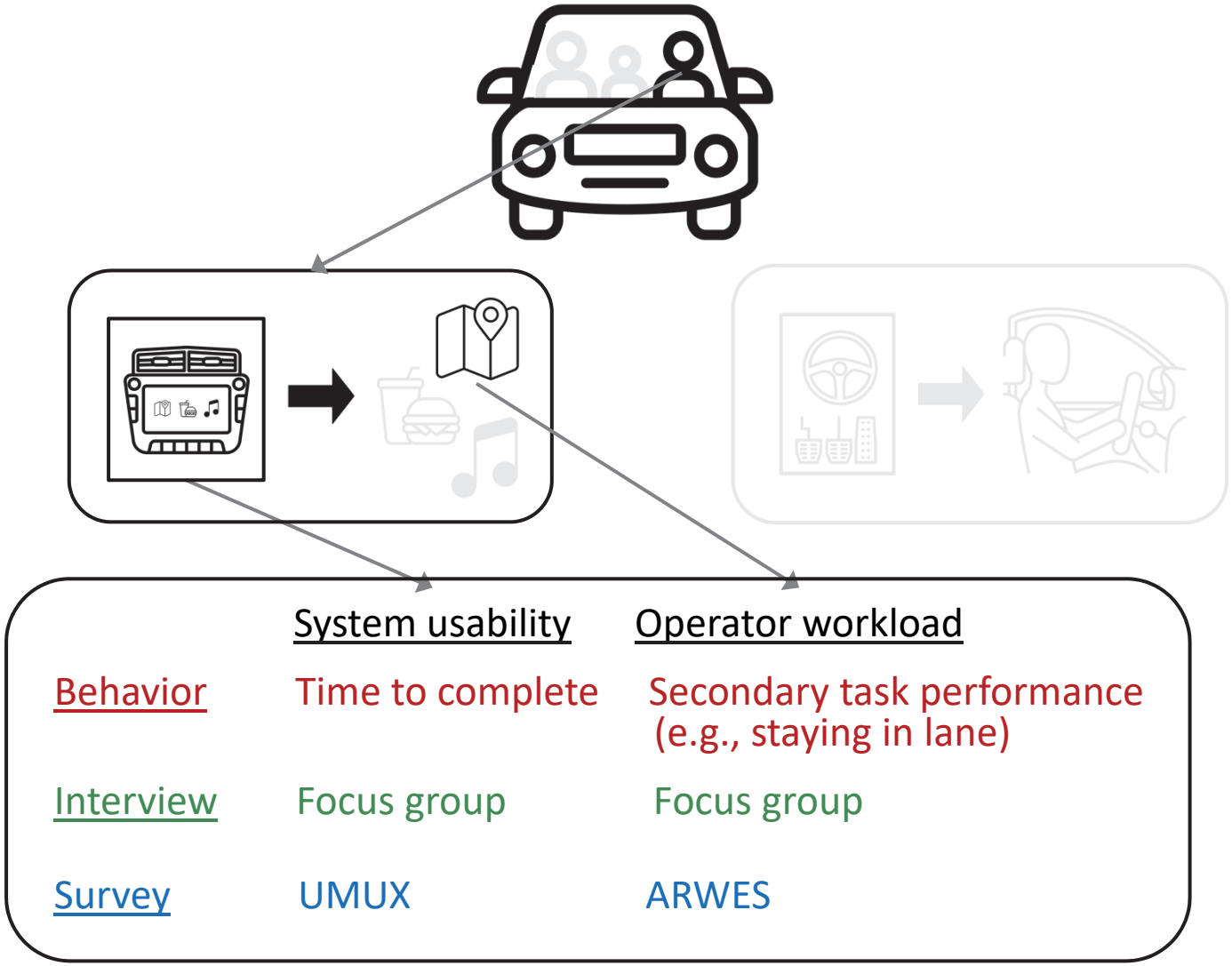
HSI – Human-Systems Interaction

Review: triangulate with HSI measurement categories



HSI – Human-Systems Interaction

Review: select appropriate HSI measures



ARWES – Air Force Flight Test Center Revised Workload Estimate Scale; HSI – Human-Systems Interaction; UMUX – Usability Metric for User Experience

Design of experiments

- Identify operational factors that might affect system performance
  - E.g., Light level, Road surface conditions
- Include these factors when planning the effectiveness DOE
- Administer HSI measures across relevant test conditions

		Light level	
		Day	Night
Road surface conditions	Dry	Behavior, Interview, Survey	Behavior, Interview, Survey
	Rain	Behavior, Interview, Survey	Behavior, Interview, Survey
	Ice	Behavior, Interview, Survey	Behavior, Interview, Survey

HSI – Human-Systems Interaction; DOE – Design of Experiments

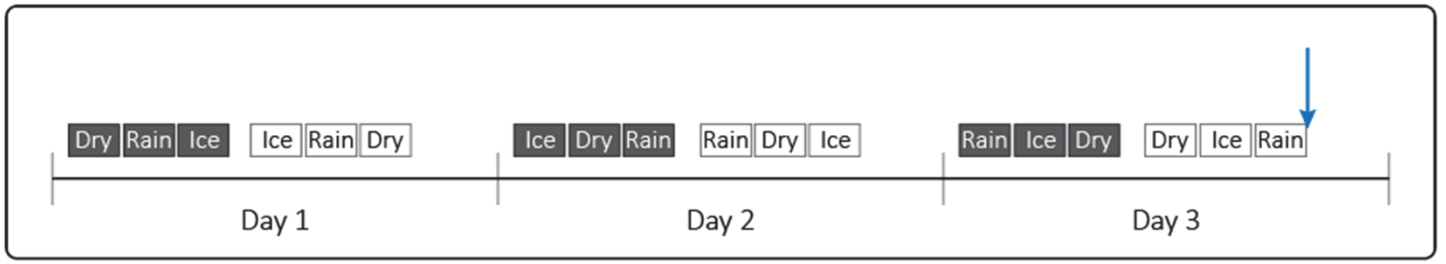
HSI administration plan: When to administer HSI measures during the test?

Administration	Typical measurement categories	Typical instrumentation	Does it support DOE analyses?
During task performance		Automated instruments	✓
After missions		Short surveys Task-focused interviews	✓
End of test		Long surveys Mission-focused interviews	✗

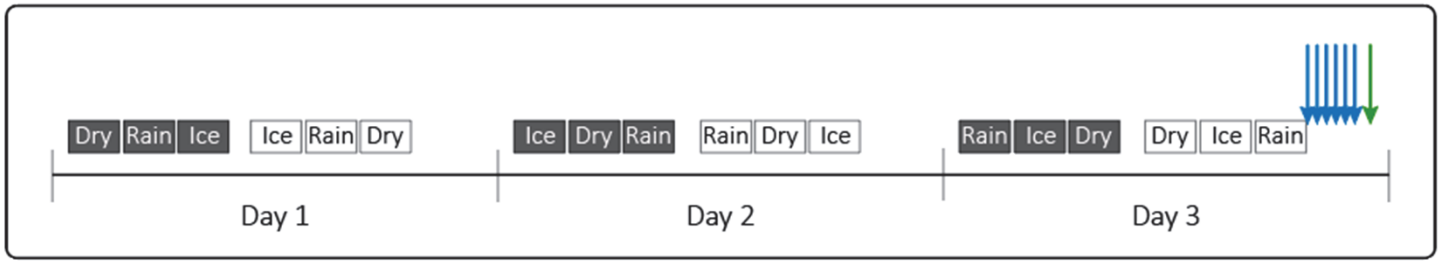
HSI – Human-Systems Interaction; DOE – Design of Experiments

Example testing schedules and administration plans

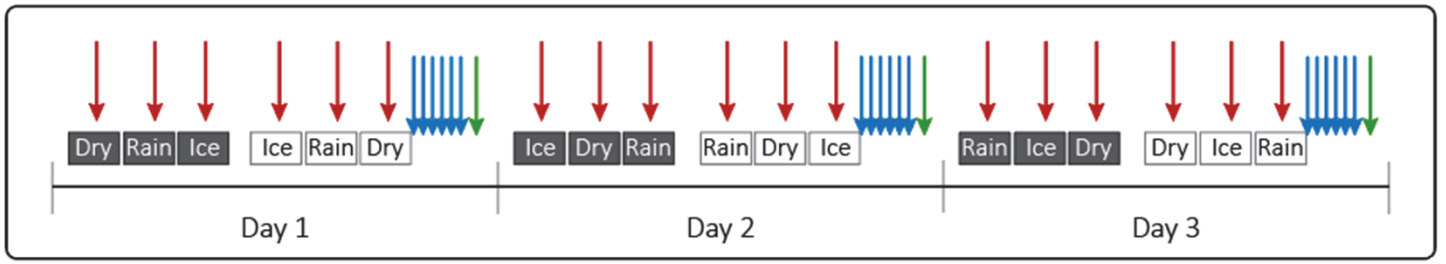
Poor



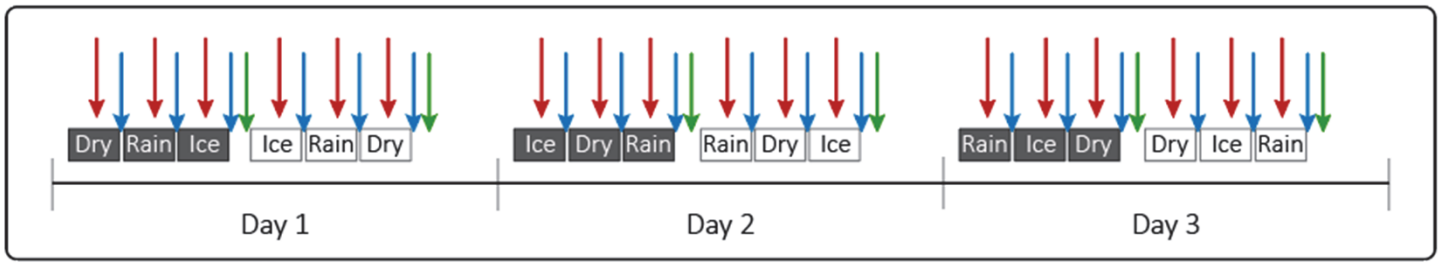
OK



Good



Great



Color key: Behavior, Interview, Survey

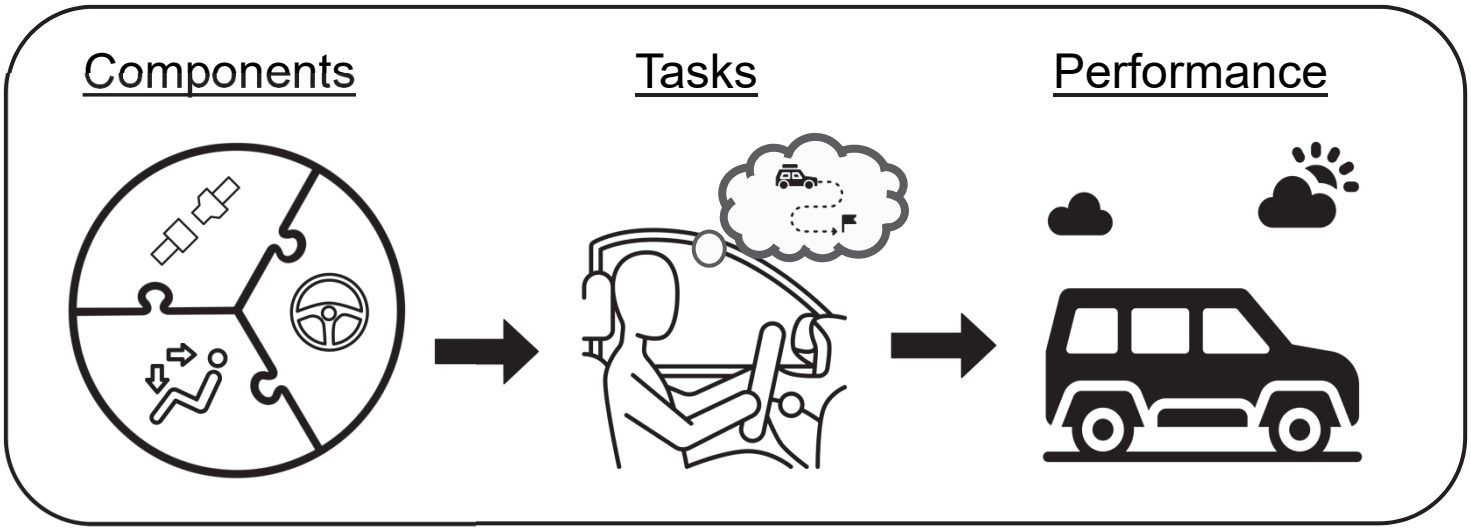
Additional consideration: operational realism

- HSI measures work best when they are administered during or immediately after an event
- However, administering HSI measures during operational events can compromise operational realism
- We recommend:
  - Administer unobtrusive HSI measures during operational events
  - If necessary, supplement data from operational events with excursions that allow intrusive HSI measures

HSI – Human-Systems Interaction

Which users to test?

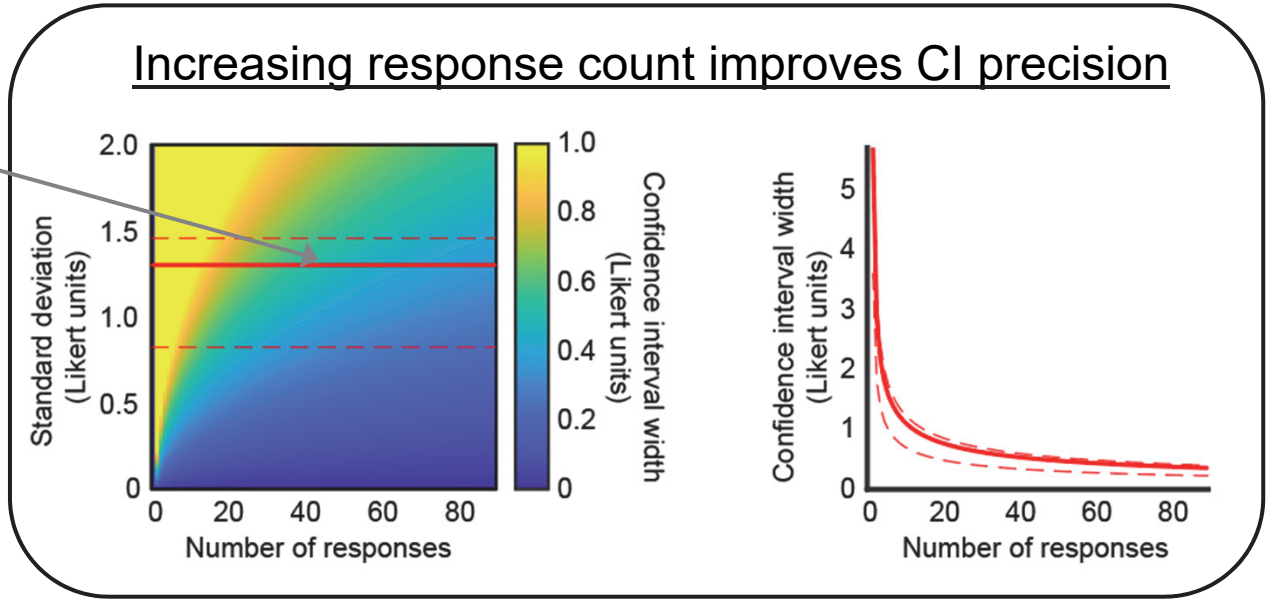
- Measure users performing tasks that affect system performance
  - Which tasks affect performance? Who performs those tasks?
- Measure representative users
  - User ability/experience can affect other HSI concepts



Sample sizes

- Sample sizes are often limited by the participants at test
  - Consider expeditions if additional data is required
- For qualitative (interview) data, aim for 10–30 participants
- For quantitative (behavior & survey) data, traditional power analyses apply
  - What is the maximum acceptable confidence interval (CI) width for parameter estimates?

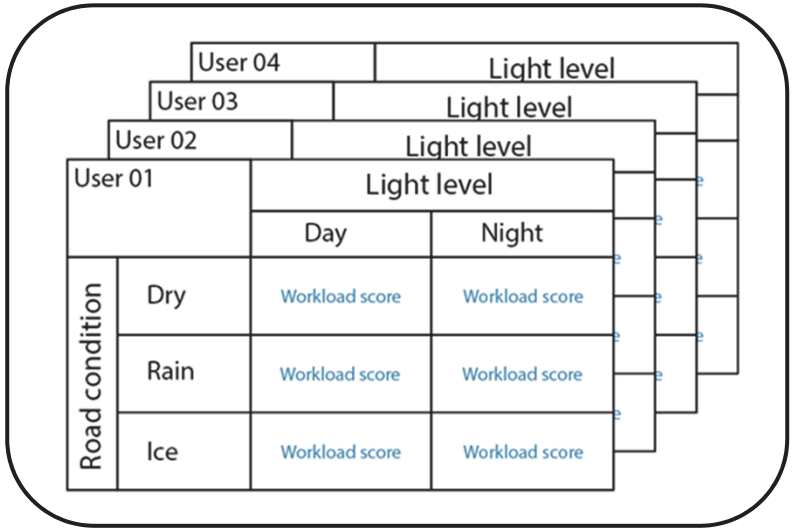
Variance of the OATS (historical)



OATS – Operational Assessment of Training Survey

Additional consideration: repeated measures

- Repeatedly measuring the same users across different conditions
  - Can be done in full (measure every user in every condition)
  - or in part (e.g., every user drives in all road conditions, but you have one group of users for minivans and another for SUVs)
- Can reduce variance and improve statistical power
- Note: may require less common power analysis methods



SUV – Sport Utility Vehicle

HSI design of experiments, key take-aways

- Plan an HSI analysis by triangulating HSI concepts that affect system performance
- Administer HSI measures across operational factors that affect system performance
- Improve your administration plan by
  - Measuring performance throughout the test (not just at the end)
  - Administering measures during or shortly after task performance
- Consider standard challenges in human research
  - Operational realism vs intrusive measurement
  - Memories fade and change with time and discussion

Contents

Analyzing HSI data

Analyzing HSI data

Analysis should mirror the design of experiments

- Identify relationships between operational factors and HSI concepts
- Capture other HSI issues, including user sentiment and HFE issues

Quantitative analysis

- Model the impact of operational factors
- Compute point estimates & confidence intervals

Qualitative analysis

- Characterize range of responses
- Identify response patterns
- Relate to HSI concepts & operational factors

HSI – Human-Systems Interaction; HFE – Human Factors Engineering

Analyzing quantitative HSI data

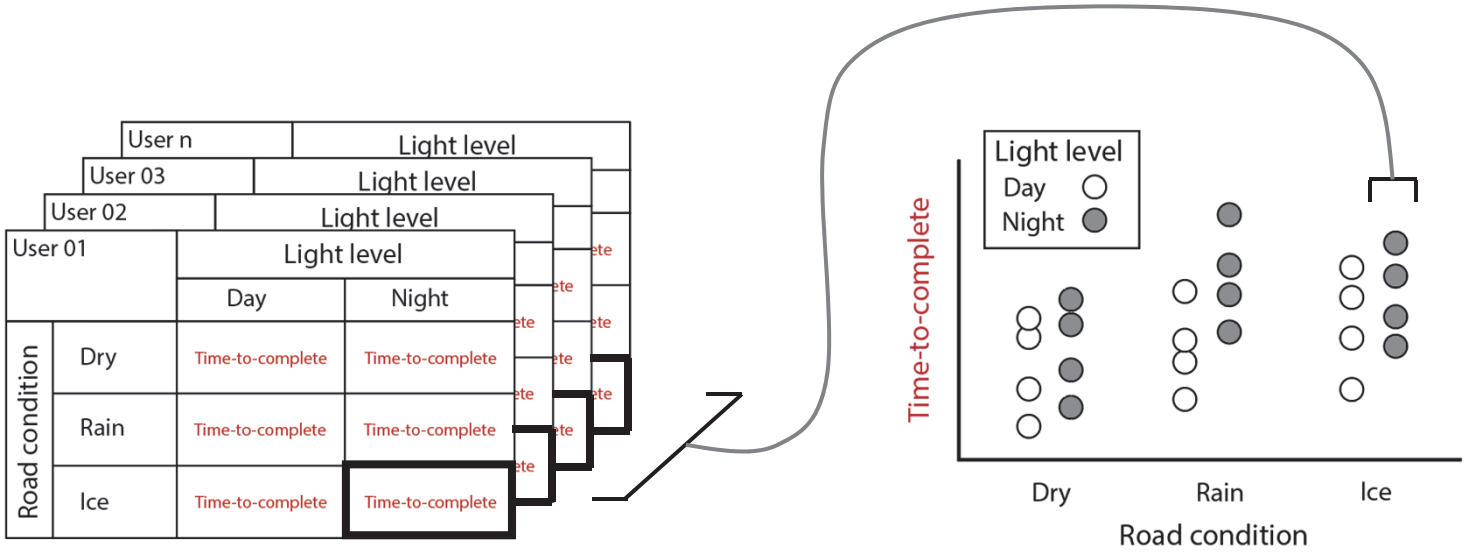
- Prepare data
  - Score surveys where appropriate
- Visualize the data
  - Plot by scenario, user, time
- Analyze sample means within operational context
  - Compute point estimates and confidence intervals
  - Compare to system requirements where relevant
- Build a model to evaluate the effects of operational conditions according to your DOE
  - Be mindful of statistical assumptions (independence, normality)
  - Designs with repeated measures require corresponding models
  - Nested data may require mixed effects models

DOE – Design of Experiments; HSI – Human-Systems Interaction

Analyzing quantitative behavioral data example

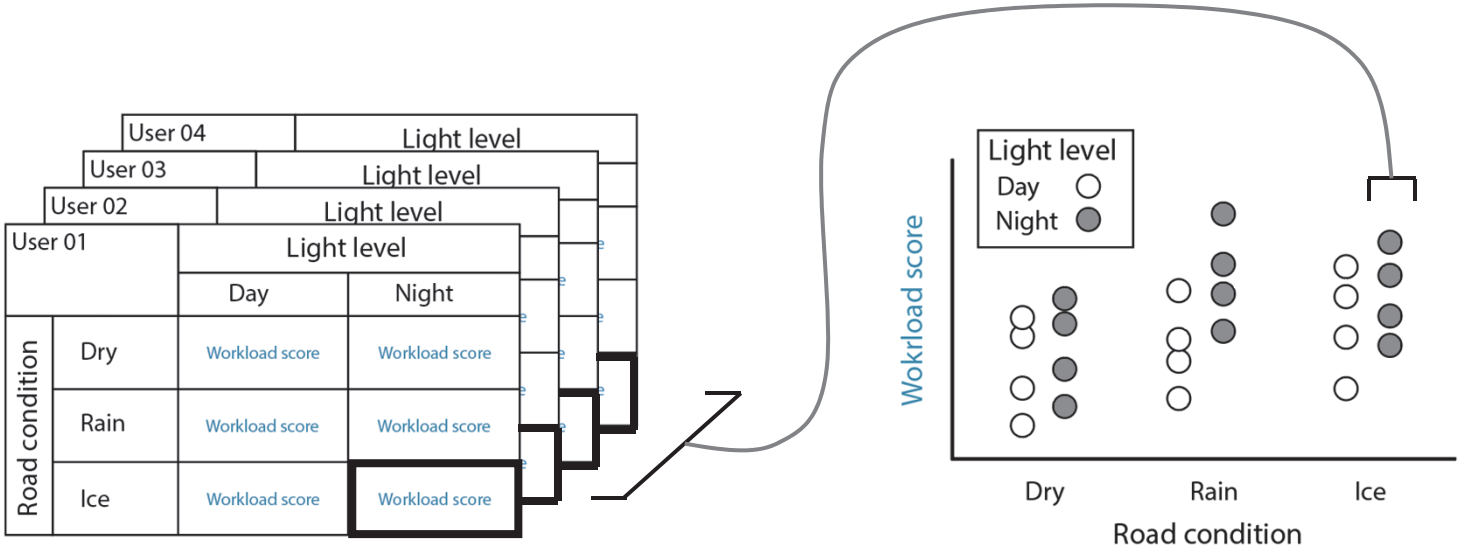
Light level

Road surface conditions	Day	Night
Dry	Behavior	Behavior
Rain	Behavior	Behavior
Ice	Behavior	Behavior



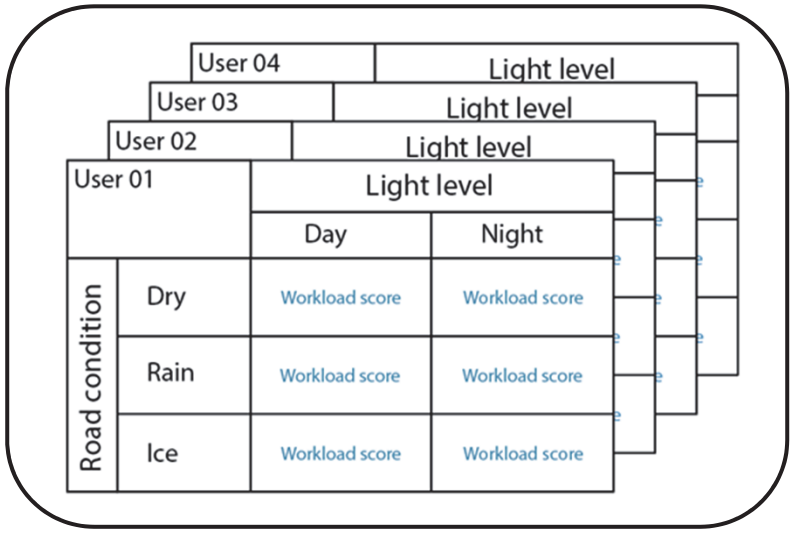
Analyzing quantitative self-report data example

Road surface conditions	Light level: Day	Light level: Night
Dry	Survey	Survey
Rain	Survey	Survey
Ice	Survey	Survey



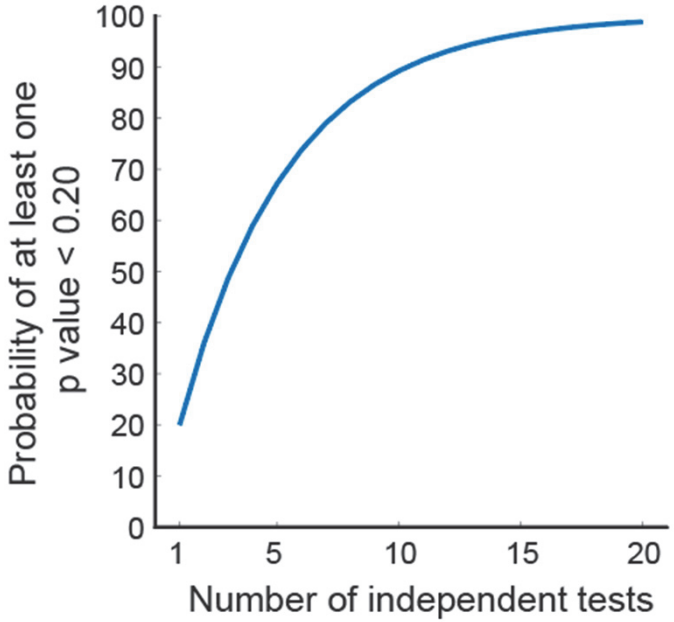
Additional consideration: repeated measures

- Do not ignore dependencies between repeated measurements
- Best practice is often to include subject ID in your model
- A more basic approach is to compute subject-level averages as a pre-processing step, leaving only one data point per subject
- Consult statistical support



Additional consideration: multiple comparisons

- Even when there are no ‘real’ effects in your data, you’re likely to find some ‘significant’ effects if you perform enough tests
- Use overarching models or apply corrections to reduce error
- Consult statistical support; practice humility



Analyzing qualitative HSI data

- Consume the data
  - Watch all of the videos
  - Read all of the user comments
- Take notes
  - What topics were discussed?
  - What ideas are repeated?
  - What are the extreme responses (i.e., the range)?
  - Record the most representative quotes
- Relate content to HSI concepts where relevant
- Note feedback that addresses impact of operational conditions

Analyzing qualitative self-report data

Focus groups

+

Open-ended survey responses

Grounded theory analysis

Theme	Repeating ideas
Value added	“Driving is way faster than walking… it’s not even close.” “We need these cars right now… gamechanger.”
Screen glare	“Touch screen buttons are useless if I can’t see them” “Difficult to see my next turn when the sun is out"
Nighttime driving	“Hard to determine the speed of approaching cars at night“ “It’s impossible to know if someone will run in front of me"
Intrusive warnings	“Seatbelt warning is very distracting… makes it more likely I will crash!"

Analyzing HSI data, key take-aways

- Build a statistical model for quantitative data that mirrors your design of experiments
- Keep qualitative data qualitative

Contents

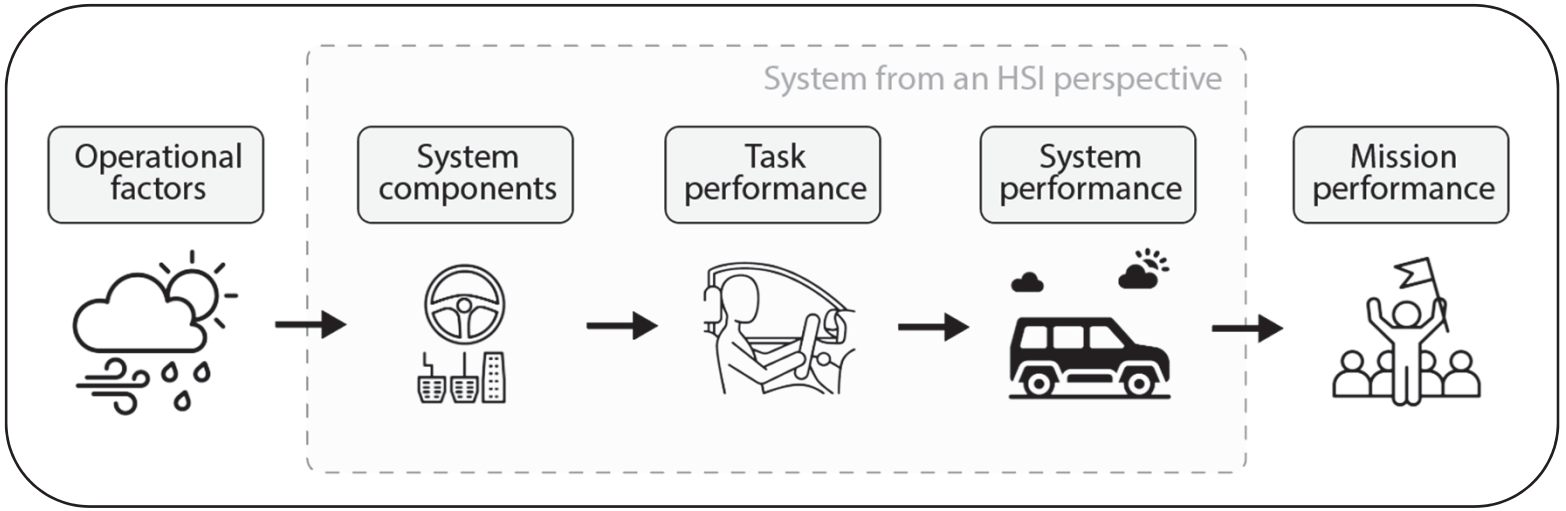
- Introduction and overview
- The system from an HSI perspective
- HSI concepts
- HSI measures
  - Validated Surveys
  - Custom Surveys
- HSI design of experiments
- Analyzing HSI data
- HSI evaluation and reporting 
- Reviewing test plans for HSI
- Resources

HSI evaluation and reporting

HSI evaluation and reporting

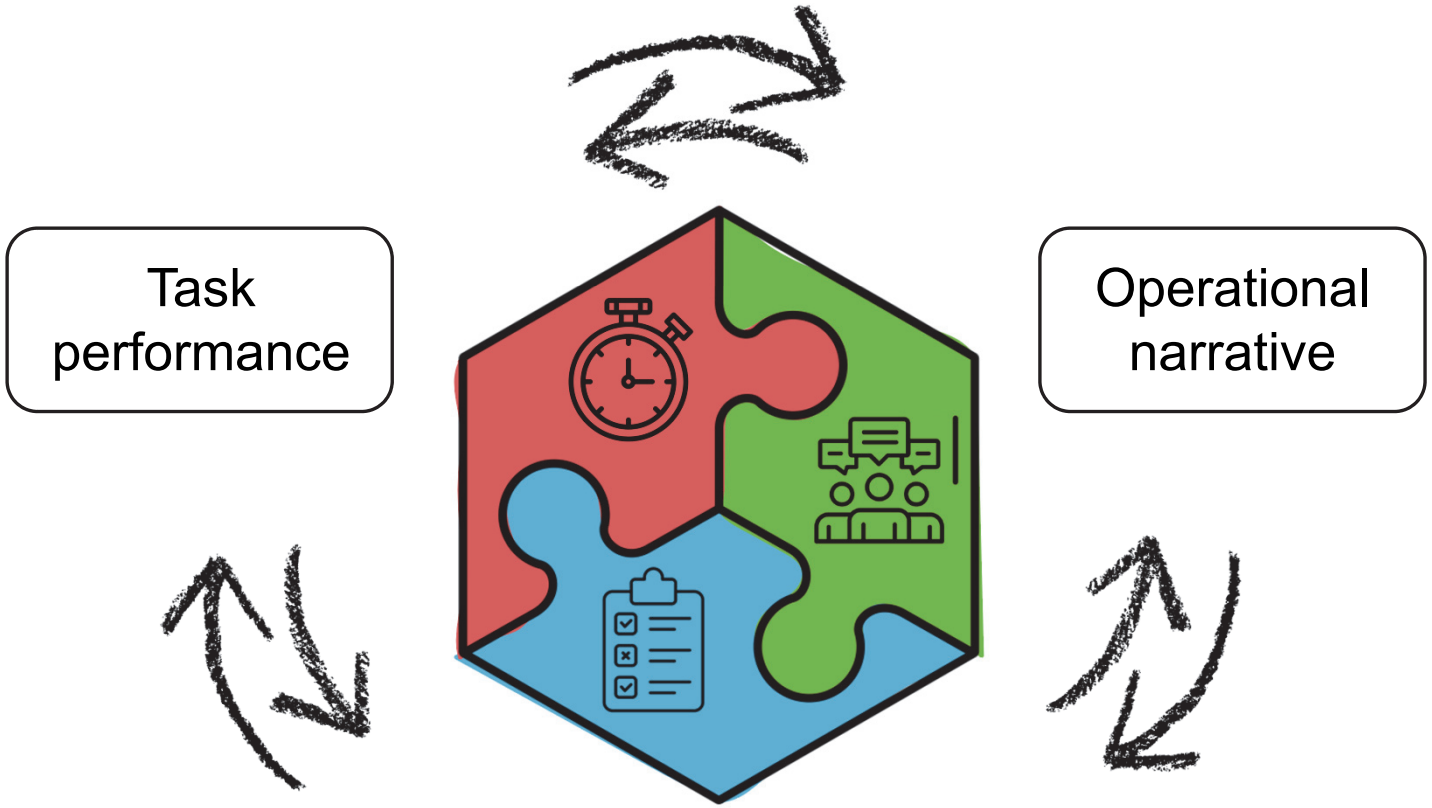
Synthesize behavior, survey, and interview data to explain *why* operational factors affect system performance

- How was HSI affected by operational conditions?
- How did HSI affect operational performance?



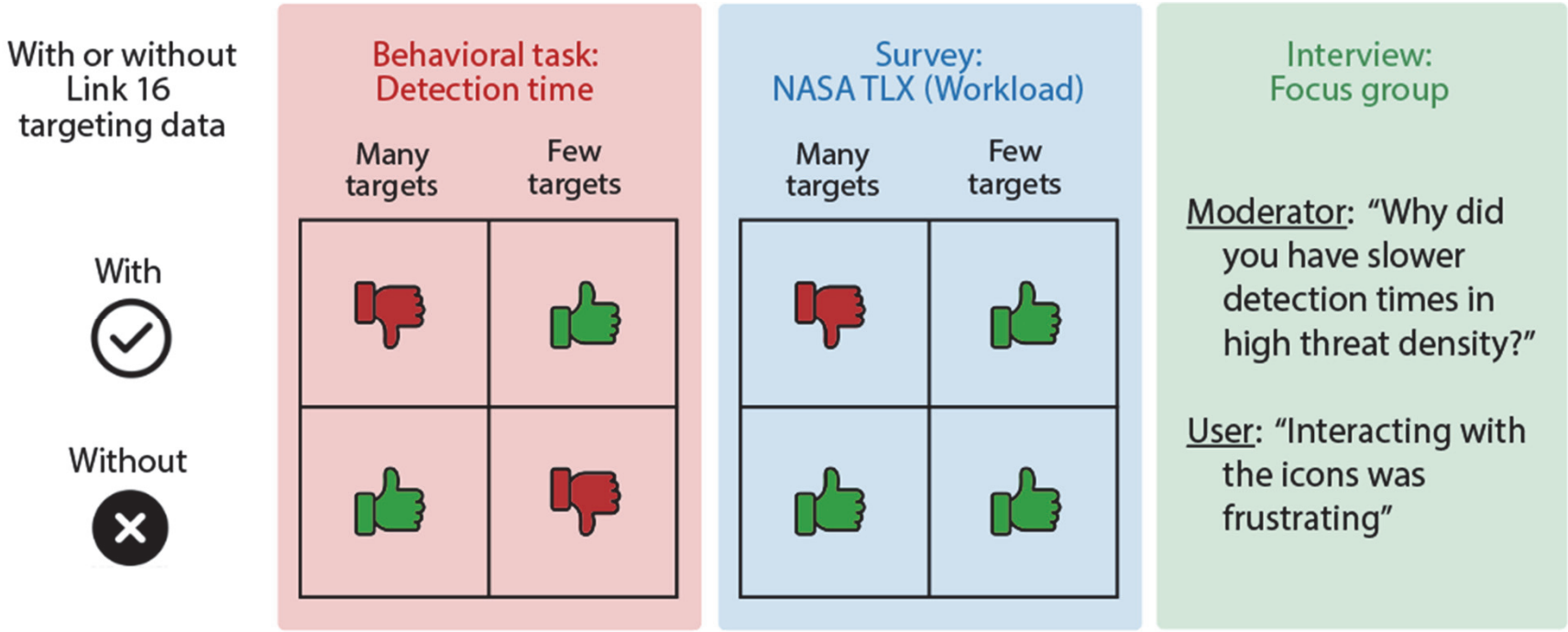
HSI – Human-Systems Interaction

The HSI evaluation is the culmination of the triangulation approach



HSI – Human-Systems Interaction

HSI evaluations synthesize HSI analyses



Conclusion: Link 16 impairs target detection when there are many targets because responding to Link 16 target notifications creates additional workload that competes with target detection task behavior

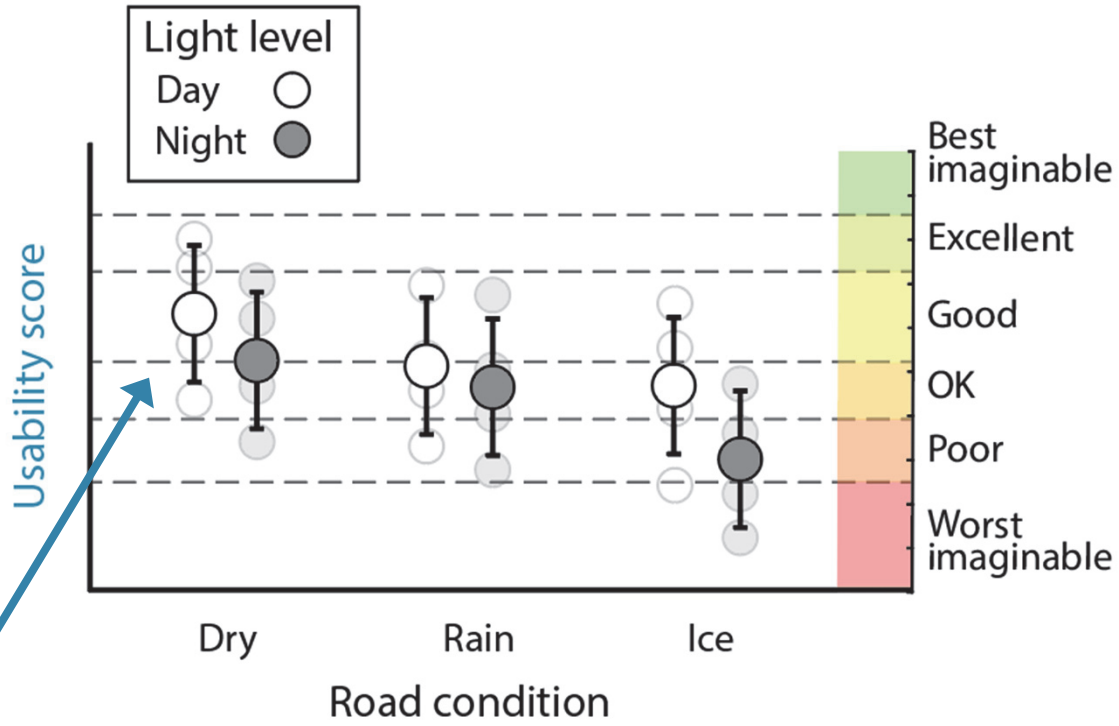
HSI – Human-Systems Interaction

Reporting quantitative HSI data

- Contextualize means and statistical test results
  - Means tell you about experiences; test results tell you about factors
  - What are the operational consequences? Reference qualitative data
- Be clear about your use of HSI data, measures, and concepts
  - Internal states are inferred, not observed directly
- Figures are the best opportunity to convey how the quantitative data support your interpretation
  - Highlight both the mean and the distribution
  - Include contextualizing information in the figure where possible

Additional consideration: Validated survey benchmarks

Use benchmark information to contextualize the effect that factors have on survey responses:



Usability is "good" in dry, daytime conditions

Reporting qualitative HSI data

- Easy to understand but difficult to convey
  - Long quotes rarely fit into briefings or reports
  - Summarizing qualitative data robs it of value
- Think about how the qualitative data informed your analysis
  - Report qualitative data to explain your quantitative choices
- Think about how the qualitative data informed your interpretation
  - Report qualitative data to contextualize the quantitative results
- Qualitative data may be the only evidence of an unexpected issue
  - Be clear about where the data ends and your interpretation begins

HSI – Human-Systems Interaction

Reporting HSI evaluations within the effective, suitable, survivable (ESS) framework

<u>Effectiveness</u> of the Joint Air-to-Ground Missile

"Effectiveness is reduced under **high pilot workloads**…due to…a **cumbersome pilot-vehicle interface** (PVI)."

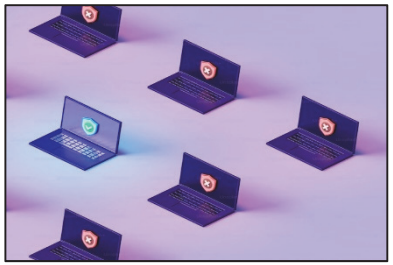
- 2022 FY DOT&E Annual Report, p 115

<u>Suitability</u> of the F-16 Radar Modernization Program

"Pilots are generally satisfied with the **human-systems interface**, although some limitations … resulted in **increased pilot workload** for some tasks."

- 2022 FY DOT&E Annual Report, p 270

Tradeoffs Between Usability and Cyber <u>Survivability</u>



Tradeoffs between (cyber)security and usability of systems mean that **poor HSI may introduce vulnerabilities** that can affect **operational survivability**.

DOT&E – Director, Operational Test and Evaluation; FY – Fiscal Year; HSI – Human-Systems Interaction

HSI evaluation and reporting, key take-aways

- The evaluation completes the triangulation by synthesizing data from different measurement categories
- Highlight narrative consistency, note inconsistency
- Keep qualitative data qualitative

Contents

Reviewing test plans for HSI

Testers must describe how they plan to implement HSI measures in the test plan. Because operators and maintainers should be able to employ the system effectively under all relevant conditions, testers should collect HSI data across these conditions.

Testers should construct a data collection plan that summarizes
(1) what HSI concepts will be measured,
(2) what method(s) they will use to measure them, and
(3) when and how they will capture the data.

The data collection plan must describe, for each test condition, the number of expected responses to HSI measures, and the number of different operators or maintainers who will provide these responses.

Guidance for Testing and Evaluating Human-System Interaction
2019 September, DOT&E memo

HSI – Human-Systems Interaction

To properly **integrate quantitative and qualitative data**, testers should:

1. Measure each of the HSI concepts with both quantitative and qualitative methods.

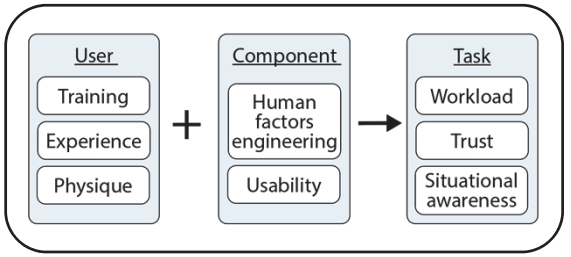
2. Integrate measures of HSI concepts into the test design so that measures of HSI can be directly linked to measured operational performance.

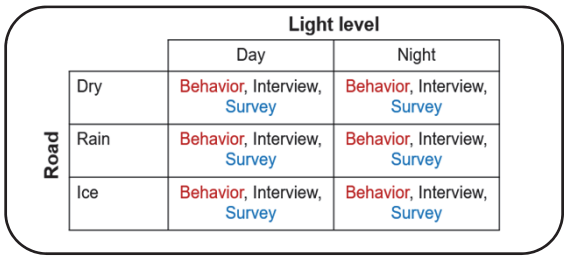
3. Collect the data in a way that permits unique operator responses to be compared with specific events in test and operational conditions.

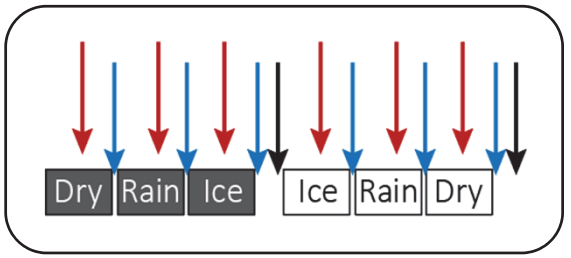
Guidance for Testing and Evaluating Human-System Interaction
2019 September, DOT&E memo

Reviewing test plans for HSI

- Test plan reviewers should consider all of the information we have discussed so far…
  - The system from an HSI perspective
  - Identifying HSI concepts and measures
  - DOE, administration, and analysis
- Reviewing HSI test plans means determining…
  - Who and what they plan to measure
  - How and when they will measure it
  - Whether their design can support the required analyses



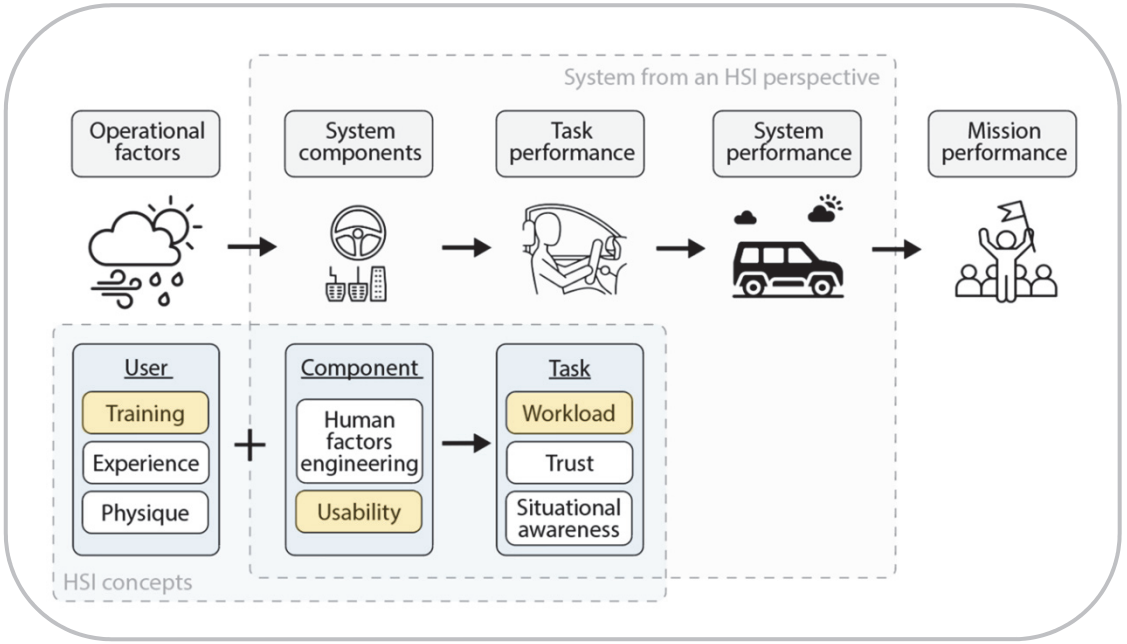




DOE – Design of Experiments; HSI – Human-Systems Interaction

Who and what do they plan to measure?

- Does their test allow you to observe HSI affecting system performance during an operational mission?
- Are they considering all of the relevant users, components, and tasks?
- Have they identified all of the relevant HSI concepts?



HSI – Human-Systems Interaction

How will they measure it?

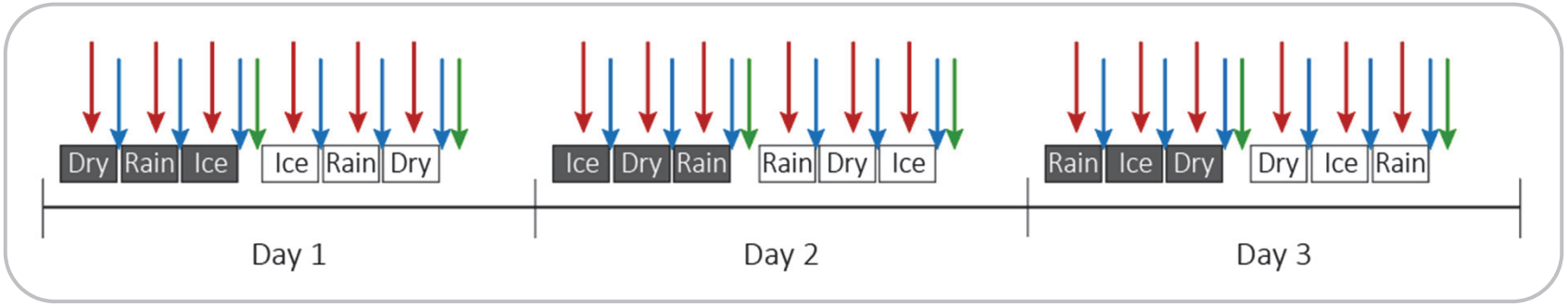
- What HSI measures will they use?
  - Are the measures feasible? Reliable? Validated? Is there a back-up plan?
- Are they triangulating all of their HSI concepts of interest?
- Will they measure the HSI concepts across the operational envelope?

		Light level	
		Day	Night
Road surface conditions	Dry	Behavior, Interview, Survey	Behavior, Interview, Survey
	Rain	Behavior, Interview, Survey	Behavior, Interview, Survey
	Ice	Behavior, Interview, Survey	Behavior, Interview, Survey

HSI – Human-Systems Interaction

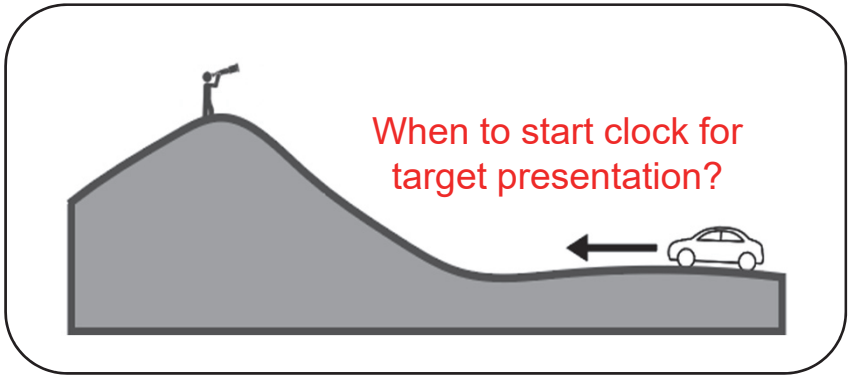
The administration plan

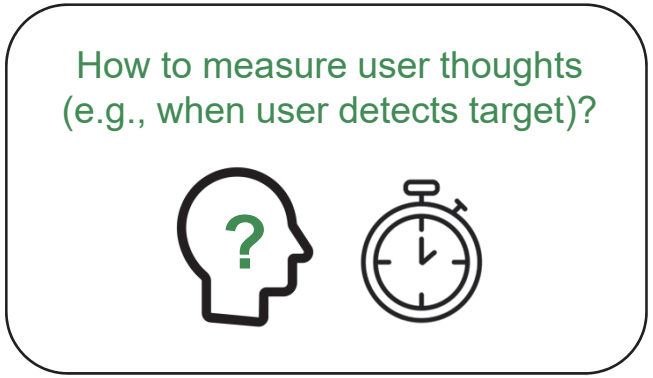
- Must specify which measures will be administered and when
- Administration tempo should balance evaluation goals against the need to preserve operational realism
- Ask for exact copies of surveys
  - Replace or improve custom surveys
  - Ensure that validated surveys are administered before custom surveys
- Ask for details about the interview and/or focus group plan
  - Are they including all of (and only) the relevant users?
  - Interviews are with 1 user; Focus groups should have around 5–10 users
  - Interviewers/moderators should be trained



Can the design support the necessary analysis?

- Many analyses fail before they begin because the test is not designed to collect the necessary data
- Look at the test design from an analyst's perspective…
  - Did they include enough participants?
  - Do they confound operational factors with learning effects?
  - Are measurements timestamped within the overall mission timeline?
  - Are unique users, stimuli, and timestamps clearly identifiable?





Reviewing test plans for HSI, key take-aways

- Check the test plan logic against everything we learned today
  - Users, components, tasks, HSI concepts, measures, DOE
- Ensure there is an administration plan
  - Characterize HSI across the operational envelope
- What might the analysis look like? Is anything missing?
- Ask for the details

DOE – Design of Experiments; HSI – Human-Systems Interaction

Contents

Resources

DoD Guidance

Most Current:

Guidance for Testing and Evaluating Human-System Interaction

2019 September, DOT&E memo

- Specifies usability, workload, and training should be measured
- Best practices for quantitative and qualitative measures

Human Systems Integration Guidebook

2022 May, USD (R&E) guidebook

- Overview of HSI considerations throughout the acquisition process

DOT&E – Director, Operational Testing and Evaluation; USD (R&E) – Undersecretary of Defense (Research & Evaluation)

IDA Resources

Evaluating HSI with AIES

2022 ATEAS Workshop: Evaluating HSI with AI-Enabled Systems: What should you consider in a TEMP? by Rachel Haga, Brian Vickers, and Daniel Porter
2022, IDA Document NS-D-33067

- HSI considerations for AI-enabled systems

Multi-Method Approach

A Multi-Method Approach to Evaluating Human-System Interactions during Operational Testing by Dean Thomas, Heather Wojton, Chad Bieber, and Daniel Porter
2017, IDA Document NS D-8857

- Additional detail on the benefits of triangulation

Introduction to Survey Design

Introduction to Survey Design by Heather Wojton, Jonathan Snavely, and Justin Mary
2016, IDA Document NS D-5835

- Slides from course covering survey methods, how to write and format surveys

The Validated Scales Repository by Test Science
https://testscience.org/validated-scales-repository/

- Lists tested survey measures and how to administer them; continually updated

User-Centric HSI

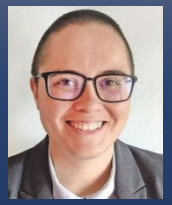
User-Centric Human Systems Integration – Evaluation Design Support by Logan K. Ausman and Daniel E. Hellmann
2022, IDA Document P-33312

- Appendixes include a mix methods (triangulation) plan design worksheet, interview guide template, and focus group moderator guide template

AI – Artificial Intelligence; ATEAS – Advancements in Test and Evaluation of Autonomous Systems; HSI – Human-Systems Interaction; TEMP – Test and Evaluation Master Plan

The Test Science HSI team at IDA is here to help

Air Warfare



Dr. Miriam Armstrong
Human Factors

PhD Human Factors Psych.
Texas Tech Univ.
@ IDA since 2022

Land Warfare

Dr. Elizabeth Green
Human Factors
Test Design

PhD Experimental Psychology
Texas Tech Univ.
@ IDA since 2021

Dr. Adam Miller
Behavioral Science
Data Analysis

PhD Behavioral Neuroscience
Cornell Univ.
@ IDA since 2022

Naval Warfare

Net Centric

Dr. Brian Vickers
AI&A, HSI
Data Science

PhD Cognitive Neuroscience
Univ Michigan, Ann Arbor
@ IDA since 2019

Dr. Sarah Shaffer
Human Factors

PhD Experimental Psych.
Florida International University
@ IDA since 2022

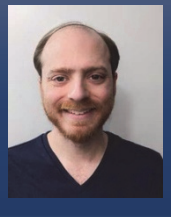
Additional expertise at IDA:

Dr. Dan Porter
AI&A
Human Factors

PhD Cognitive Neuroscience
Univ. of Michigan
@ IDA since 2017

Ms. Rachel Haga
CDAO Lead,
AI&A, Human Factors

MS Aerospace Engineering
Georgia Tech
@ IDA since 2020

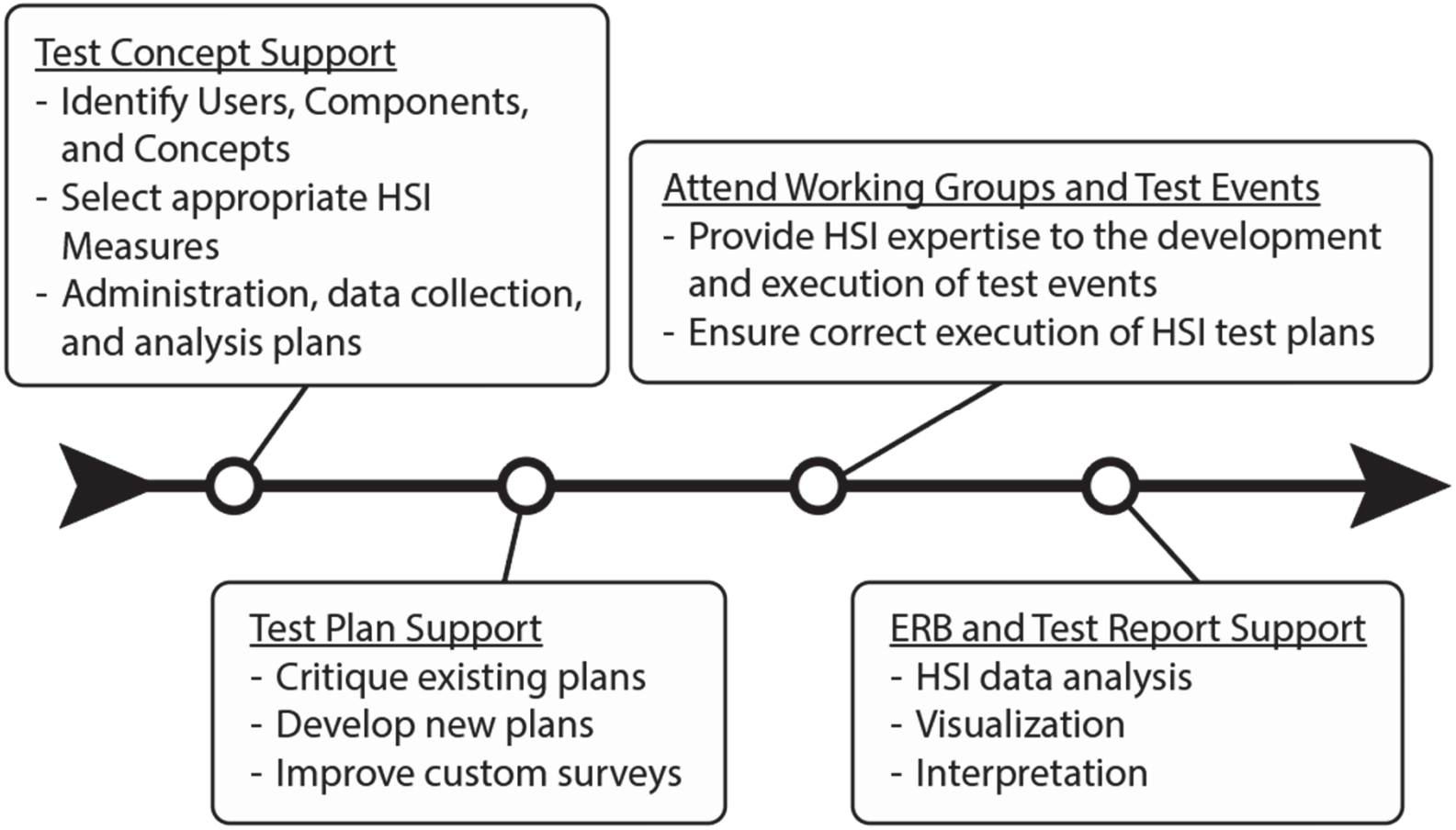


Mr. Yosef Razin
Human Factors,
Robotics, Trust

BS Mechanical Engineering
Georgia Tech
@ IDA since 2022

AI&A – Artificial Intelligence and Automation; CAP – Cyber Assessment Program; CDAO – Chief Digital and Artificial Intelligence Office; HSI – Human-Systems Interaction

Test Science HSI support to programs



ERB – Emerging Results Brief; HSI – Human-Systems Interaction

Backups

Look-up table for selecting appropriate HSI methods

Table 1. General Guidance for When to Apply Common HSI Measurement Methods

	Quantitative			Qualitative			
Test Team Goals	Behavior	Validated Survey	Custom Survey	Interview	Focus Group	Comment Boxes	Testers' Qualitative Observation
1. Measure operational performance	X						
2. Quantify effect of HSI on operational performance	X	X	X				
3. Quantify recognized HSI concept		X					
4. Describe operator experiences				X	X	X	
5. Problem discovery or diagnosis				X	X	X	X
6. Manage resource constraints		X			X	X	
7. Get high quality data	X	X		X			
8. Reduce operator burden	X	X		X	X		X

Guidance for Testing and Evaluating Human-System Interaction

2019 September, DOT&E memo

DOT&E – Director, Operational Test & Evaluation; HSI – Human-Systems Interaction

Validated surveys

Training: Operational Assessment of Training Scale (OATS)

- Perceived educational value of training
- Two subscales: Relevance & Efficacy
- Administer twice: after NET and after test

Item		Strongly Disagree						Strongly Agree
1	All of the information covered was relevant to how I interact with the system.	1	2	3	4	5	6	7
2	The training prepared me to easily use the system to accomplish my mission.	1	2	3	4	5	6	7
3	Training accurately portrayed operations in the field.	1	2	3	4	5	6	7
4	The training prepared me to properly interact with the system.	1	2	3	4	5	6	7
5	Training prepared me to solve common problems.	1	2	3	4	5	6	7
6	Training adequately covered all important ways I interact with the system.	1	2	3	4	5	6	7

NET – New Equipment Training

Training: Diagnostic Survey of Testing (DSoT)

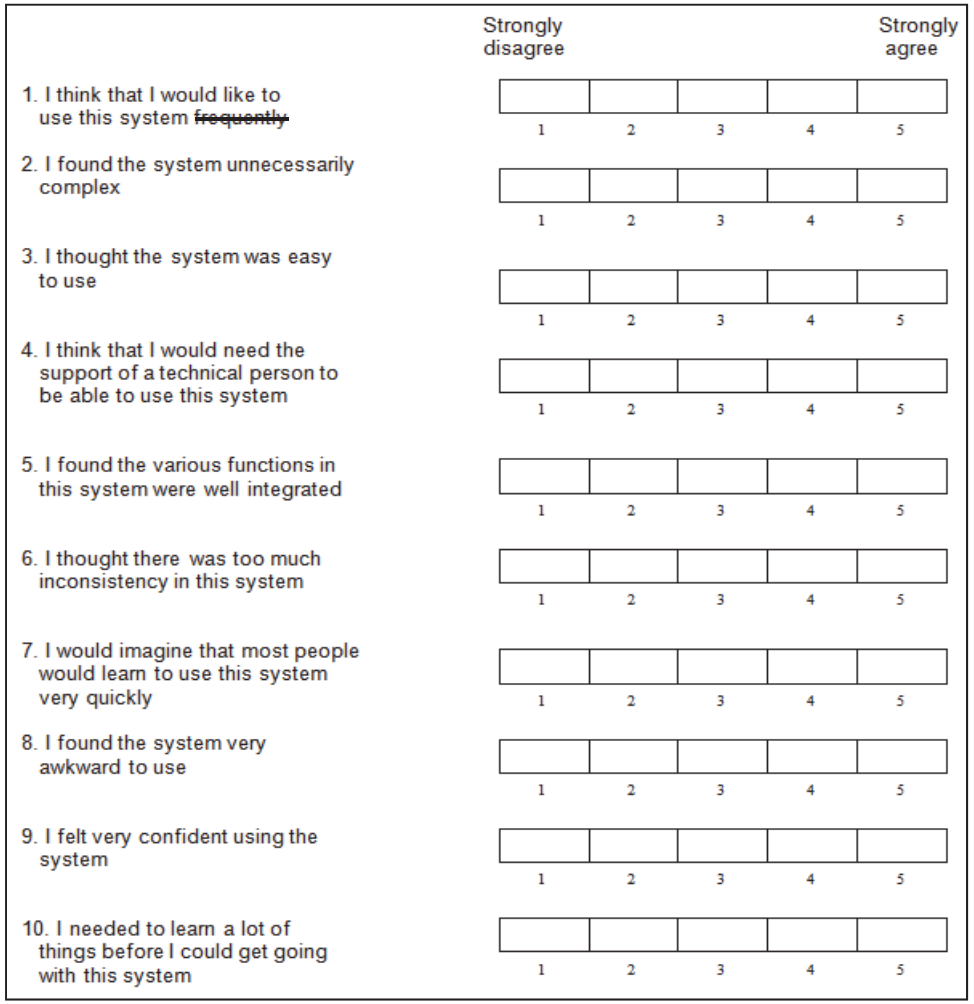
- Diagnostic tool
- How would you change the training?
- Also known as the Training Structure Assessment (TSA)

Item		Significantly Decrease			No Change			Significantly Increase
1	Amount of hands-on training	1	2	3	4	5	6	7
2	Amount of lecture	1	2	3	4	5	6	7
3	Detail of course training content	1	2	3	4	5	6	7
4	Pace of the course training	1	2	3	4	5	6	7
5	Amount of reference materials provided	1	2	3	4	5	6	7
6	Amount of time for questions	1	2	3	4	5	6	7
7	Reinforcement of course training content	1	2	3	4	5	6	7

Item		Strongly Disagree						Strongly Agree
1	The instructor did a good job overall	1	2	3	4	5	6	7

Usability: System Usability Scale (SUS)

- Easy-to-interpret scoring guide available
- Multiple versions exist
  - All positive (recommended)
  - Edited wording



	Strongly disagree				Strongly agree
1. I think that I would like to use this system ~~frequently~~	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very awkward to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Usability: Usability Metric for User Experience Lite (UMUX-Lite)

- Benchmark to SUS scoring
- Less sensitive than SUS
- Good for repeated administration
- A four-item version is also available (UMUX)

Item		Strongly Disagree						Strongly Agree
1	[This system's] capabilities meet my requirements.	1	2	3	4	5	6	7
2	[This system] is easy to use.	1	2	3	4	5	6	7

SUS – System Usability Scale

Workload: NASA Task Load Index (NASA TLX)

- Multiple types of workload
- Benchmarks available
- Standard administration is a little complicated

- There is an approved 'raw' version with simplified administration

Item	Factor		Response Line	
1	Mental Demand	Low	|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|	High
2	Physical Demand	Low	|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|	High
3	Temporal Demand	Low	|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|	High
4	Performance	Good	|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|	Poor
5	Effort	Low	|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|	High

Workload: AFFTC Revised Workload Estimate Scale (ARWES)

- Easy interpretation
- Good for repeated administration
- Only appropriate for mental workload

Statement	
1	Nothing to do; No system demands.
2	Light Activity; minimal demands.
3	Moderate activity; easily managed considerable spare time.
4	Busy; Challenging but manageable; Adequate time available.
5	Very busy; Demanding to manage; Barely enough time.
6	Extremely busy; Very difficult; Non-essential tasks postponed.
7	Overloaded; System unmanageable; Essential tasks undone; Unsafe.

AFFTC – Air Force Flight Test Center

Trust: Trust of Automated Systems Test (TOAST)

- Not only for automated systems
- Two subscales: Understanding and Performance

Item		Strongly Disagree						Strongly Agree
1	I understand what the system should do.	1	2	3	4	5	6	7
2	The system helps me achieve my goals.	1	2	3	4	5	6	7
3	I understand the limitations of the system.	1	2	3	4	5	6	7
4	I understand the capabilities of the system.	1	2	3	4	5	6	7
5	The system performs consistently.	1	2	3	4	5	6	7
6	The system performs the way it should.	1	2	3	4	5	6	7
7	I feel comfortable relying on the information provided by the system.	1	2	3	4	5	6	7
8	I understand how the system executes tasks.	1	2	3	4	5	6	7
9	I am rarely surprised by how the system responds.	1	2	3	4	5	6	7

Image citations

Image citations, 1 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	HSI effectiveness image	7, 109	Samuel Ruiz	https://www.dvidshub.net/image/7076692/marine-light-attack-helicopter-squadron-267-equipping-joint-air-ground-missile-marine-corps-air-station-camp-pendleton	Public Domain
	HSI suitability image	7, 109	Audra Flanagan	https://www.148fw.ang.af.mil/News/Photos/igphoto/2003136146/	Public Domain
	Purple laptops	7, 109	Rodion Kutsaiev	https://unsplash.com/photos/a-group-of-purple-laptops-sitting-on-top-of-each-other--mCL702J-lA	https://unsplash.com/plus/license
	Passenger jet icon	8	Uswa KDT	https://thenounproject.com/icon/plane-6991099/	CCBY
	Three jets icon	8	WEBTECHOPS LLP	https://thenounproject.com/icon/jet-1694581/	CCBY
	Rocket icon	8	Vectorstall	https://thenounproject.com/icon/missile-6608825/	CCBY
	Helicopter icon	8	Ehtisham Abid	https://thenounproject.com/icon/helicopter-7016858/	CCBY
	User on laptop icon	8	Support Designs	https://thenounproject.com/icon/laptop-user-1680462/	CCBY
	Radar screen	8	Suprihatin	https://thenounproject.com/icon/radar-6992547/	CCBY

HSI – Human-Systems Interaction; CCBY – Creative Commons Attribution License

Image citations, 2 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	Thought cloud icon	8	Bakunetsu Kaito	https://thenounproject.com/icon/thinking-918417/	CCBY
	Radar image	8	Kmg Design	https://thenounproject.com/icon/radar-6924547/	CCBY
	Stopwatch icon	9, 40-44, 53, 80, 83, 92-94, 104, 106, 115, 117, 119	DinosoftLabs	https://thenounproject.com/icon/stopwatch-5231971/	CCBY
	Survey icon	9, 40, 41, 45, 50-53, 58-63, 80, 83, 92, 93, 95, 98, 99, 104, 106, 107, 115, 117, 130-136	Timo Schmid	https://thenounproject.com/icon/survey-6293386/	CCBY
	Interview icon	9, 40, 41, 45-49, 53, 80, 83, 92, 98, 104, 108, 115, 117	ProSymbols	https://thenounproject.com/icon/focus-group-1875637/	CCBY
	Car icon	10, 16, 19, 20, 24-27, 33, 35, 42, 78, 79, 86, 103, 116	Design Circle	https://thenounproject.com/icon/pick-up-4288089/	CCBY
	Circular puzzle	16, 78, 86	Libby Ventura	https://thenounproject.com/icon/three-piece-circle-puzzle-outline-2440186/	CCBY
	Seatbelt icon	16, 78, 86	throwaway icons	https://thenounproject.com/icon/security-belt-343102/	CCBY
	Airflow icon	16, 78, 86	LAFS	https://thenounproject.com/icon/ac-1747449/	CCBY
	Steering wheel icon	16-18, 20, 25-27, 33, 35, 78, 79, 81, 86, 103, 116	Alena Artemova	https://thenounproject.com/icon/wheel-2061818/	CCBY

CCBY – Creative Commons Attribution License

Image citations, 3 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	Driver icon	16, 18, 19, 20, 25, 26, 27, 33, 35, 42, 78, 79, 81, 86, 103, 116	Amethyst Studio	https://thenounproject.com/icon/driving-5221006/	CCBY
	Car travel icon	16, 18, 78, 86	Slamet Widodo	https://thenounproject.com/icon/road-trip-6143000/	CCBY
	Driving pedals icon	17, 18, 20, 25-27, 33, 35, 79, 81, 103, 116	Designing Hub	https://thenounproject.com/icon/pedals-6199909/	CCBY
	Infotainment icon	17, 18, 31, 81	Icongeek26	https://www.flaticon.com/free-icon/car-audio_3902923	https://www.flaticon.com/legal#nav-flaticon-agreement
	Map icon	17, 18, 20, 31, 81	Secondtoughest	https://thenounproject.com/icon/map-6659537/	CCBY
	Food icon	17, 18, 20, 31, 81	Satria Arnata	https://thenounproject.com/icon/food-6638762/	CCBY
	Music icon	17, 18, 20, 31, 81	Kiki Rizky	https://thenounproject.com/icon/music-6659254/	CCBY
	Car from front icon	17, 18, 81	Designing Hub	https://thenounproject.com/icon/family-car-6298779/	CCBY
	Coffee cup	17	Nur Hasanah	https://thenounproject.com/icon/coffee-cup-5427689/	CCBY
	Baby icon	18	Tomasz Pasternak	https://thenounproject.com/icon/baby-2032155/	CCBY
Image	Image description	Slide appearance(s)	Author	Link to source	License

CCBY – Creative Commons Attribution License

Image citations, 4 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	Victory icon	20, 24-27, 33, 35, 79, 103, 116	Iconatrenda	https://thenounproject.com/icon/victory-6588483/	CCBY
	Javelin NET image 1	28	Lisa M. Sadler	https://www.flickr.com/photos/vaguardpao/49385263733	https://creativecommons.org/licenses/by-nc/2.0/
	Javelin NET image 2	28	Tad Browning	https://www.army.mil/article/269303/cavalry_troopers_test_new_javelin_missile_system	Public Domain
	Weather icon	24-27, 33, 35, 79, 103, 116	Hanbai	https://thenounproject.com/icon/weather-6795622/	CCBY
	Roundabout image	32	WIRED	https://www.wired.com/2016/08/brilliant-sorcery-englands-7-circle-magic-roundabout/	CC BY-NC
	Experience icon	34	Bahrul Ulum	https://thenounproject.com/icon/experience-5975840/	CCBY
	Awareness icon	34	Ivan Colic	https://thenounproject.com/icon/awareness-30176/	CCBY
	Trust icon	34	Icon 5	https://thenounproject.com/icon/trust-6224517/	CCBY
	Black box icon	36	Fithratul Hafizd	https://thenounproject.com/icon/box-icon-3699540/	CCBY
	Stethoscope icon	36	SBTS	https://thenounproject.com/icon/stethescope-2476313/	CCBY

CCBY – Creative Commons Attribution License; CC BY-NC – Creative Commons Attribution Non-Commercial License

Image citations, 5 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	Javelin image	43	Tad Browning	https://www.dote.osd.mil/Portals/97/pub/reports/FY2023/army/2023javelin.pdf	Public Domain
	Head icon	44, 47, 52, 53, 119	Adrian Coquet	https://thenounproject.com/icon/head-1912200/	CCBY
	Focus group icon	46	Iconathon	https://thenounproject.com/icon/meeting-2909311/	CCBY
	One-on-one icon	46	Creative Studio	https://thenounproject.com/icon/interview-6750693/	CCBY
	Pen-and-paper icon	46, 99	Mada Creative	https://thenounproject.com/icon/write-5999570/	CCBY
	Flower and eggs icon	58, 67	Vectors Point	https://thenounproject.com/icon/food-ingredients-3186048/	CCBY
	Mixer icon	58	Mar'atus Ratna Kurni	https://thenounproject.com/icon/mix-ingredients-5648392/	CCBY
	Cake icon	58	Win Ningsih	https://thenounproject.com/icon/cake-6687399/	CCBY
	Checkmark icon	83, 105	NAPISAH	https://thenounproject.com/icon/check-4728432/	CCBY
	X icon	83, 105	Royyan Wijaya	https://thenounproject.com/icon/x-3922739/	CCBY

CCBY – Creative Commons Attribution License

Image citations, 6 of 6

Image	Image description	Slide appearance(s)	Author	Link to source	License
	Movie icon	99	Sandhi Priyasmoro	https://thenounproject.com/icon/video-1547462/	CCBY
	Black chalk arrow	104	Spencer Harrison	https://thenounproject.com/icon/curved-arrow-4120281/	CCBY
	Thumb icon	105	Ilaria Bernareggi	https://thenounproject.com/icon/thumbs-up-734176/	CCBY

CCBY – Creative Commons Attribution License

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
XX-06-2024	Final	

4. TITLE AND SUBTITLE

Introduction to Human Systems Interaction in Operational Test and Evaluation Course

5a. CONTRACT NUMBER

HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Adam M. Miller

5d. PROJECT NUMBER

BD-9-2299(98)

5e. TASK NUMBER

229998

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
730 East Glebe Road
Alexandria, Virginia 22305

8. PERFORMING ORGANIZATION REPORT NUMBER

3002009

3002009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)

DOT&E

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

This OED Draft has not been approved by the sponsor for distribution and release. Reproduction or use of this material is not authorized without prior permission from the responsible IDA Division Director.

13. SUPPLEMENTARY NOTES

Project Leader: Keyla Pagan-Rivera

14. ABSTRACT

Human-System Interaction (HSI) is the study of interfaces between humans and technical systems. The Department of Defense incorporates HSI evaluations into defense acquisition to improve system performance and reduce lifecycle costs. During operational test and evaluation, HSI evaluations characterize how a system's operational performance is affected by its users. The goal of this course is to provide the theoretical background and practical tools necessary to plan and evaluate HSI test plans, collect and analyze HSI data, and report on HSI results. We will discuss HSI concepts, measurement methods, design of experiments, data analysis, and evaluation and reporting, all from an operational testing perspective.

15. SUBJECT TERMS

Keywords: HSI, human-systems, human factors, operational, DOE, analysis, suitability, usability, workload, training, trust, situational awareness, survey, behavior, interview, focus group, qualitative, quantitative, mixed-methods, triangulation, validated scale

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Keyla Pagan-Rivera
Unclassified	Unclassified	Unclassified	SAR	154	19b. TELEPHONE NUMBER *(Include area code)* 703-845-6936