NASA-STD-3001 Technical Brief

# Usability, Workload, & Error

OCHMO-TB-005
Rev B

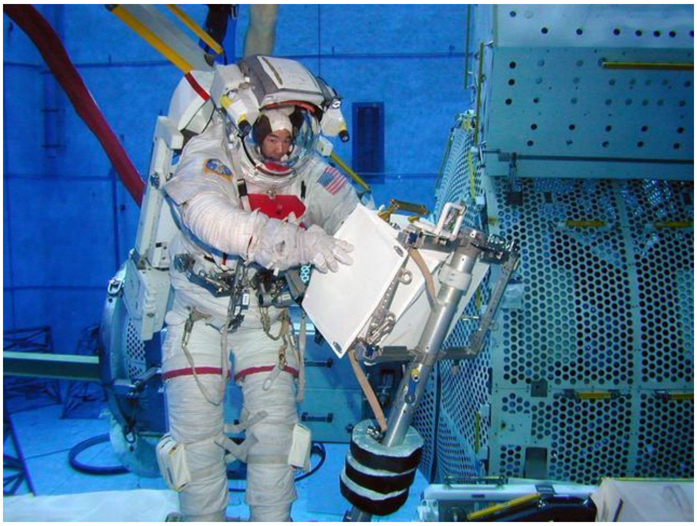
## Relevant Technical Requirements

**NASA-STD-3001 Volume 1, Rev B**
[V1 4014] Completion of Critical Tasks

**NASA-STD-3001 Volume 2, Rev C**
[V2 3006] Human-Centered Task Analysis
[V2 5004] Cognitive Capabilities
[V2 5007] Cognitive Workload
[V2 10001] Crew Interface Usability
[V2 10002] Design-Induced Error
[V2 10003] Crew Interface Operability
[V2 10200] Physical Workload

## Executive Summary

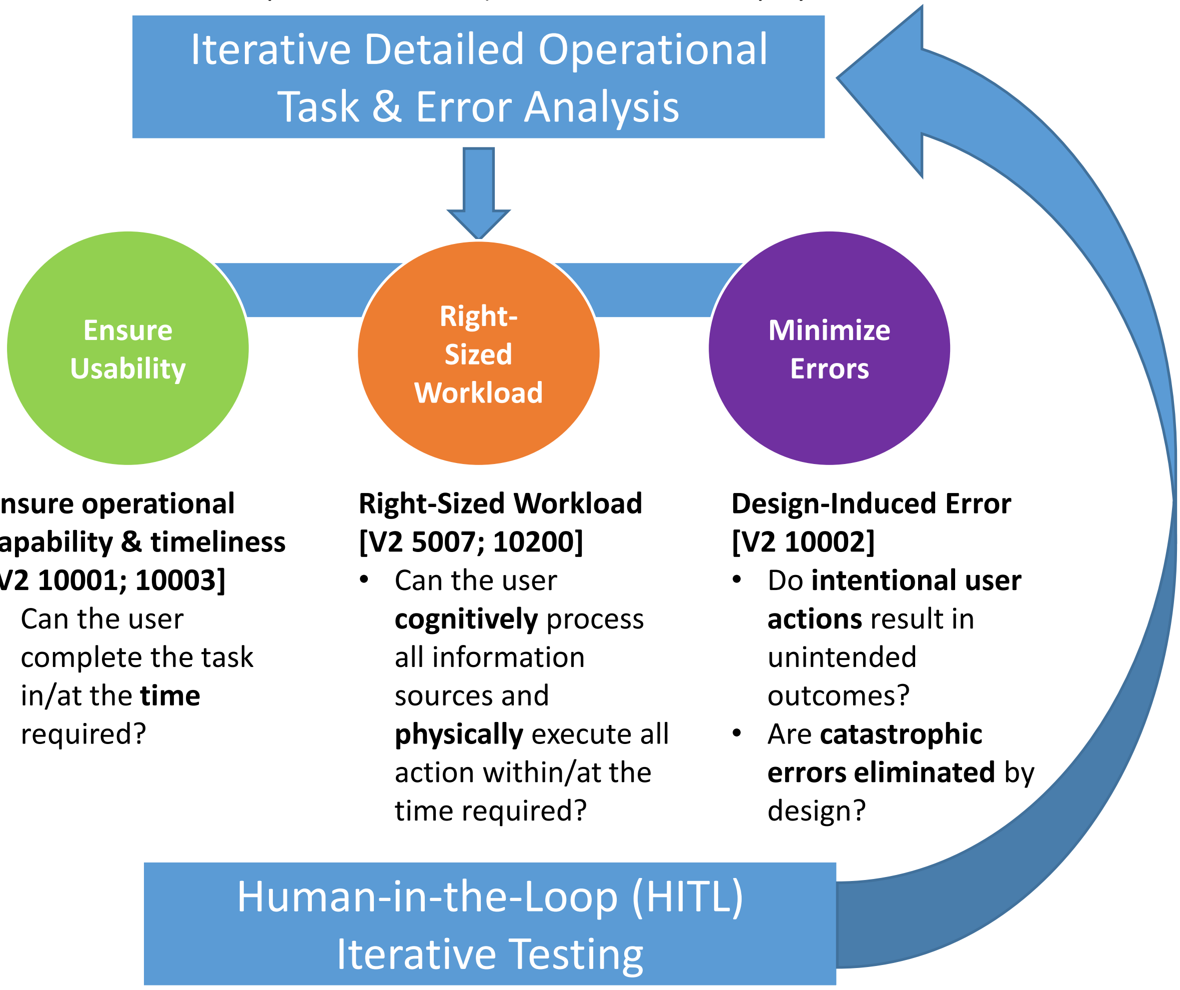

An iterative design approach that considers critical and complex tasks the crew must perform should be thoroughly tested with human-in-the-loop testing. The goal is to minimize errors, ensure crew usability and right size workload to enable crew performance/mission success. Early and iterative testing in the design cycle will identify issues to minimize requirement non-compliance and design changes later in the cycle when they are more expensive to implement and can cause significant schedule impacts.

*Crewmember testing hardware usability in NASA's Neutral Buoyancy Lab*

# Background

Generating a detailed concept of operations and describing the tasks the crew is expected to perform is critical to ensuring the proper design of the vehicle. Design considerations include layout of crew displays and controls, concurrent activities and the ability of the crew to perform within the given volume with access to required resources (such as switches, displays, tools, latches, hatches, etc.).




**Ensure operational capability & timeliness [V2 10001; 10003]**

- Can the user complete the task in/at the **time** required?

**Right-Sized Workload [V2 5007; 10200]**

- Can the user **cognitively** process all information sources and **physically** execute all action within/at the time required?

**Design-Induced Error [V2 10002]**

- Do **intentional user actions** result in unintended outcomes?
- Are **catastrophic errors eliminated** by design?

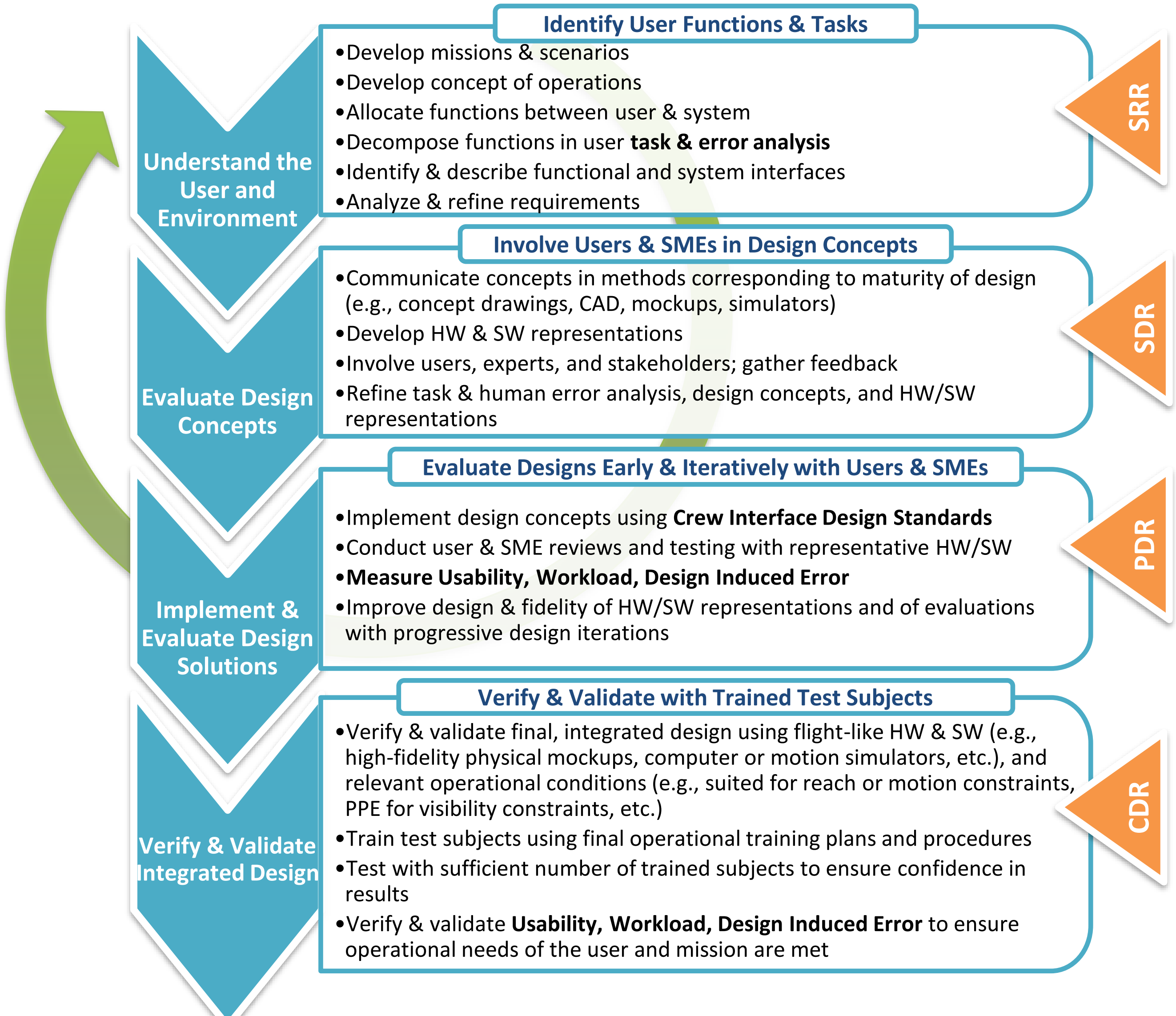
The goal is to have an effective and efficient system developed through operational task analysis and user testing that supports Total Mission Performance.

**Risk of Inadequate Operational Task Analysis and User Testing**

Skylab 4: While preparing for entry, crew inadvertently opened a circuit break for the wrong control system. Unaware of the erroneous switch position, crew were unable to command the vehicle to proper attitude for re-entry. Crew switched to manual backup to save the vehicle and crew. Poor switch interface design (location and labeling) contributed to design-induced error during a critical and intense task. Task and error analysis should have identified error potential so that design controls could have been implemented. Manual backup control allowed crew to save the mission when automation become ineffective (G. Johnson, JSC-2018-009).

# Reference Data

## User-Centered Design Process




Systems that are usable are acceptable and operable by the intended user for performing expected tasks. If a design does not meet the users' needs, expectations, intuitions, or capabilities, and as a result causes frustration or confusion, the design is not effective.

**Risk of Inadequate Design for Usability**

The usability of the design of systems and equipment on the space station Mir raised moderate levels of concern amongst the crew. Design related comments comprised 11.2% of all total comments for Mir, 40.0% of which were negative in nature. Crew comments and lessons learned have shown NASA that inadequate space design methodologies continue to be an issue for ISS (Baggerman, Rando, & Duvall).

# Reference Data

## Design-Induced Error

Human error analysis identifies potential user errors at each step and the outcome or system consequence if the error is committed. Task errors that would result in a catastrophic outcome should be prevented through careful interface design and thorough evaluation. Tasks that are identified as complex, leading to critical or catastrophic hazards/events, or frequent need more rigorous developmental testing and are to be included in HITL verification testing.

For purposes of HITL testing, a scenario requiring evaluation will be defined as an activity driven by one or more related and sequential procedures. The procedure consists of a series of task steps, where a task step will be defined as a single instruction to the test subject, as is typical of current space flight procedures. Participants will maintain task completion times commensurate with the performance requirements.

If any errors classified as having the potential of leading to a catastrophic outcome occur, the root cause of the error must be identified, mitigated satisfactorily (approved by NASA), and a re-test of the task performed to prove that the error has been eliminated.

The percentage of errors (erroneous task steps) for each user shall be calculated by dividing the number of erroneous task steps and incomplete task steps by the total number of task steps and multiplying the result by 100.

The percentage of users committing each error (erroneous task step) shall be calculated by dividing the number of users committing each erroneous task step by the total number of users and multiplying the result by 100.

**NASA-STD-3001 [V2 10002] Design-Induced Error** requires system designs to provide crew interfaces that result in the maximum observed error rates listed in Table 29 (below).

**Risk of Inadequate Testing for Design-Induced Error**
Errors are detrimental to crew effectiveness, efficiency, acceptability, and safety. Even when recoverable or resulting in minimal impact, errors can still negatively impact crew performance in terms of productivity and satisfaction.

Errors are defined as an action that does not result in the intended outcome or a failure by the crew to perform an action within the required limits of accuracy, sequence, or time which results in unwanted consequences. Design-induced errors include, but are not limited to: missed or incorrect inputs or selections, display navigation errors, errors due to inadequate hardware component design, errors due to lack of system feedback to user inputs, errors due to inadequate information, errors due to design inconsistency or unfamiliar terminology, and the inability to complete a step or task.

**Table 29 – Maximum Observed Design-Induced Error Rates**

| Type of Error | Maximum Observed Error Rate |
|---|---|
| Catastrophic Error | 0% |
| Non-Catastrophic Errors per User per Task | 5% |
| Non-Catastrophic Errors per Step per Task | 10% |

# Application

**Decompose concept of operations into function allocation and operational task and error analysis early in the design**

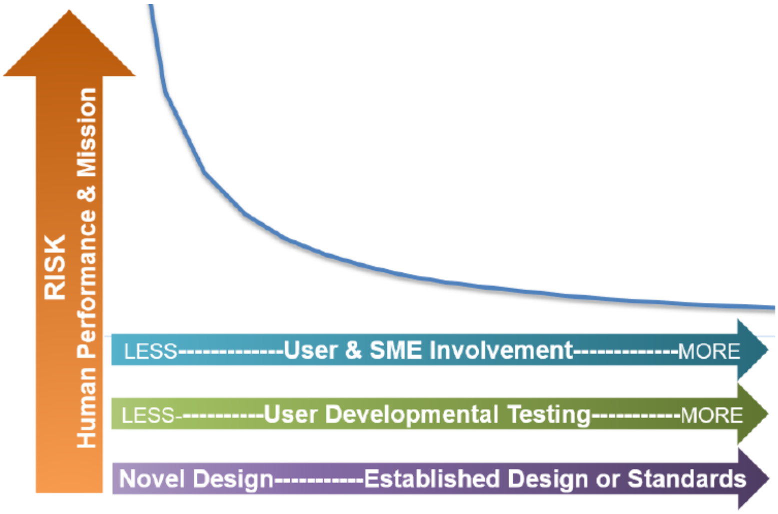
- Identify nominal and contingency operations
- Identify automated functions and manual override capabilities
- Identify critical functions and potential for human error
- Update mission concept of operations/function allocation/task and error analysis as design matures throughout the development cycle.

**Involve users and Subject Matter Experts (SMEs) to perform early developmental testing and evaluation**

- Testing with representative users and SMEs helps to evolve and refine design
- Simulator testing is beneficial to identify inadvertent operation error potential
- Minimizes redesign and unplanned costs later in the development cycle
- Reduces operation risk, especially for novel or non-standard designs

**Verify and validate final integrated design**

- Use flight-like hardware and software (e.g., high-fidelity physical mockups, computer or model simulators, etc.)
- Use validated operational procedures
- Simulate relevant operational conditions (e.g., suited for reach or motion constraints, personal protective equipment for visibility constraints, etc.)
- Test with sufficient number of trained subjects to ensure confidence in results. See the Human-in-the-Loop (HITL) Test Subject Sample Sizes OCHMO Technical Brief for additional information.




**NASA-STD-3001 [V2 10001] Crew Interface Usability** requires system designs to provide crew interfaces that result in a NASA-Modified System Usability Scale (NMSUS) score of 85 or higher. See page 6 of this technical brief for the full NMSUS. *From NASA-STD-3001 Volume 2, Rev C*

# Back-Up

## Development and Validation of the NASA Modified System Usability Scale (NMSUS): A Brief Summary

Usability (the ability of a user to complete a task with effectiveness, efficiency, and satisfaction) is often assessed during spacecraft verification testing, along with other human performance metrics (e.g., workload, errors). Over a recent one- to two-year period of spacecraft verification tests, human factors subject matter experts noted that on multiple occasions, astronauts had comments regarding the face validity of the System Usability Scale (SUS; Brook, 1996*) that was being used to measure usability. The SUS is a widely used and accepted measure of perceived usability within the human factors community, and is supported by a robust literature. Although the SUS scale is well validated and has been used to measure perceived usability for a wide variety of products and systems, crew comments indicated that the phrasing of some of the scale items did not align with the types of safety-critical tasks being completed as part of spacecraft verification.

A Human Research Program (HRP) project was undertaken to explore ways to address crew comments regarding the SUS, with the broader goal of improving verification testing for all future space programs. The output of this project was a tailored SUS, shorter, and better aligned with NASA tasks and terminology: the NASA Modified SUS (NMSUS; see figure on next page). Changes were minor and consisted of rephrasing three items and removing two items.

### Validation of the NMSUS

A study was conducted to assess and validate the NMSUS in terms of internal reliability (how strongly a set of items within a measure relate to one another); convergent validity (the extent to which two measures of the same target construct agree); sensitivity (the ability of a measure to reliably detect differences between interfaces); and equivalence (high similarity between SUS and NMSUS results). In the validation study, 35 crew-like participants completed procedures-driven tasks related to configuring a hypothetical backup electrical power system on a spacecraft. They interacted with two different prototypes to complete the tasks (one “well-designed”, and one “less-well-designed”). In each case, they completed a post-test survey consisting of the SUS, NMSUS, and the Usability Metric for User Experience (UMUX).

Results demonstrated that the NMSUS was as reliable as the SUS. Likewise, results supported the validity of the NMSUS in terms of convergent validity with the UMUX. The NMSUS was also determined to be sensitive, as there was a statistically significant difference in perceived usability as measured by the NMSUS between the “well-designed and the less-well-designed prototypes. Finally, multiple significance tests showed the NMSUS and SUS scores to be “practically equivalent”.

Taken together, the tests indicate that the NMSUS can be used with confidence in lieu of the SUS for measuring perceived usability at NASA. The scale should be considered a suitable replacement for the SUS when used in the safety-critical spaceflight domain. The NMSUS is currently part of NASA requirements sets for multiple spaceflight programs and scheduled for inclusion in NASA-STD-3001.

# NASA-Modified System Usability Scale (NMSUS)

The NMSUS is a validated usability scale that was developed to better align with NASA tasks and terminology. It is a tailored version of the System Usability Scale (SUS), with fewer items and some modified phrasing. See https://measuringu.com/sus/ for more information on the original SUS.

NASA Modified System Usability Scale (SUS) - NMSUS

| | | Strongly Disagree<br>1 | 2 | 3 | 4 | Strongly Agree<br>5 |
|---|---|---|---|---|---|---|
| | 1. I thought the system was easy to use. | i | i | i | i | i |
| | 2. I think that I would need technical support to be able to use this system. | i | i | i | i | i |
| | 3. I found the various functions in this system were well integrated. | i | i | i | i | i |
| | 4. I thought there was too much inconsistency in this system. | i | i | i | i | i |
| | 5. I imagine that most trained crewmembers would learn to use this system very quickly. | i | i | i | i | i |
| | 6. I found the system very cumbersome to use. | i | i | i | i | i |
| | 7. I felt very confident using the system. | i | i | i | i | i |
| | 8. I needed a lot of training on this system in order to get going. | i | i | i | i | i |

**Scoring**

The NMSUS is scored using the following equation:

SUS = 3.125 * ((Q1-1)+(5-Q2)+(Q3-1)+(5-Q4)+(Q5-1)+(5-Q6)+(Q7-1)+(5-Q8))

See also Lewis & Sauro, *The Factor Structure of the System Usability Scale*

# Major Changes Between Revisions

Rev A → Rev B

- Added information on the NASA-Modified System Usability Scale (NMSUS)
- Added information on Design-Induced Error

Original → Rev A

- Updated information to be consistent with NASA-STD-3001 Volume 1 Rev B and Volume 2 Rev C.
- Added risk of inadequate usability example.

# Referenced Technical Requirements

**NASA-STD-3001 Volume 1 Revision B**
**[V1 4014] Completion of Critical Tasks** The planned number of hours for completion of critical tasks and events, workday, and planned sleep period shall have established limits to assure continued crew health and safety.

**NASA-STD-3001 Volume 2 Revision C**
**[V2 3006] Human-Centered Task Analysis** Each human space flight program or project shall perform a human-centered task analysis to support systems and operations design.

**[V2 5004] Cognitive Capabilities** The system shall accommodate anticipated levels of crew cognitive capabilities under expected tasks demands.

**[V2 5007] Cognitive Workload** The system shall provide crew interfaces that result in Bedford Workload Scale ratings of 3 or less for nominal tasks and 6 or less for off-nominal tasks.

**[V2 10001] Crew Interface Usability** The system shall provide crew interfaces that result in a NASA-modified System Usability Scale (SUS) score of 85 or higher.

**[V2 10002] Design-Induced Error** The system shall provide crew interfaces that result in the maximum observed error rates listed in Table 29, Maximum Observed Design-Induced Error Rates.

**[V2 10003] Crew Interface Operability** The system shall provide interfaces that enable crewmembers to successfully perform tasks within the appropriate timeframe and degree of accuracy.

**[V2 10200] Physical Workload** The system shall provide crew interfaces that result in a Borg-CR10 rating of perceived exertion (RPE) of 4 (somewhat strong) or less.

## Reference List

1. Johnson, G. (2018). Gary Johnson: Lessons Learned from 50+ Years in Human Spaceflight and Safety. JSC-2018-009. https://ntrs.nasa.gov/citations/20190028301
2. Hirshorn, S.R., Voss, L.D., & Bromley, L.K. (2017). NASA Systems Engineering Handbook. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170001761.pdf
3. Zumbado, J.R. (2015). Human Systems Integration (HSI) Practitioner's Guide. https://ntrs.nasa.gov/citations/20150022283
4. NASA Systems Engineering Processes and Requirements (w/Change 1). NPR 7123.1C. https://nodis3.gsfc.nasa.gov/displayDir.cfm?t=NPR&c=7123&s=1B
5. Baggerman, S.D., Rando, C.M., & Duvall, L.E. Habitability and Human Factors: Lessons Learned in Long Duration Space Flight. *American Institute of Aeronautics and Astronautics*. https://www.nasa.gov/sites/default/files/atoms/files/medical_habitability_and_human_factors.pdf
6. James R. Lewis and Jeff Sauro. *The Factor Structure of the System Usability Scale*. Retrieved from: https://measuringu.com/wp-content/uploads/2017/07/Lewis_Sauro_HCII2009.pdf
7. Brooke, J. (1996). SUS - A quick and dirty usability scale. Retrieved from: https://digital.ahrq.gov/sites/default/files/docs/survey/systemusabilityscale%2528sus%2529_comp%255B1%255D.pdf
8. Lewis, J.R., and Sauro, J. (2009). The Factor Structure of the System Usability Scale. *International Conference on Human Centered Design.*