Initial Validation of the Trust of Automated Systems Test (TOAST)

Dean Thomas, Project Leader
Heather M. Wojton
Stephanie T. Lane
Daniel J. Porter

March 2018
Approved for public release. Distribution is unlimited.
IDA Non-Standard Document
NS D-8996
Log: H 2018-000097

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

Approved for public release; distribution is unlimited.
INSTITUTE FOR DEFENSE ANALYSES

IDA Non-Standard Document NS D-8996

Initial Validation of the Trust of Automated Systems Test (TOAST)

Dean Thomas, Project Leader
Heather M. Wojton
Stephanie T. Lane
Daniel J. Porter
Approved for public release; distribution is unlimited.
Initial Validation of the Trust of Automated Systems Test (TOAST)

Heather Wojton,
Stephanie Lane,
Daniel Porter

03/21/2018
Trust is the expectation that a person or system will help an operator achieve his goal in vulnerable situations.
Missile Warning Systems
$r \approx 0.71$

Trust $\rightarrow$ Reliance

(Muir, 1989)
Mistrust Puts the Operator and the Mission at Risk

Disuse OR Misuse

Under-trust

Over-trust
Goal: validate a scale to measure trust in automated systems
Theory points to 3 factors driving trust in automation

Purpose → Performance → Processes

What you’re told
Designer’s Intent

Your experience
Others’ experiences

Your expertise
IDA designed a scale to represent these factors

Trust in automated systems

Purpose

"I understand what the system should do."

Performance

"The system performs as it should."

Processes

"I understand how the system executes tasks."
Establishing validity
Approved for public release; distribution is unlimited.
Establish the number of factors that the scale measures

**Confirmatory factor analysis** tests how well a hypothesized structure fits the observed data.

Scale quality is determined for each subscale.

Goodness of fit Indices: $\chi^2, CFI, and RMSEA$
We expected items to cluster into 3 factors: purpose, performance, and process
Establish that the subscales are correlated with theoretically related concepts

We expected subscales to positively correlate with the statement “I trust the system”
Study 1
Approved for public release; distribution is unlimited.
Cadets were randomly assigned to read vignettes

Imagine that you are conducting a search and rescue mission aided by an unmanned ground vehicle called UGV S350. The vehicle is designed to move independently through a conflict zone, scanning buildings to determine if victims are trapped inside. When a victim is detected, the vehicle alerts the rescue unit by sounding an alarm. During the mission, UGV S350…

High Reliability
successfully alerted the rescue unit to the presence of both civilian and military personnel that were trapped during the conflict and aided in their recovery. All rescued parties were transported to a local hospital for treatment.

Low Reliability
falsely alerted the rescue unit that a victim was trapped inside a building, and the rescue unit was unable to locate a victim after performing a comprehensive search. This extended the duration of the mission and placed the rescue unit at risk.
Cadets completed the reliance intentions questions

“If placed in a similar situation in real life, I would rely on the UGV S350 to locate victims during search and rescue missions.”

“How reliable is the system?”

Not at all reliable | | | | | | Completely reliable
Manipulation Check Successful!
Cadets completed the TOAST & validation question

350 Cadets

UGV S350 Vignette → Reliance Intentions

TOAST

Trust Question

“I trust the system.”
How many factors are there?
We fit a 1-, 2-, and 3-factor model to the data

Purpose, Performance, Process
The 1- and 3-factor models were poor fits

\[ \chi^2(74) = 387.53, p < .01, CFI = .79, RMSEA = .11 \]
A 2-factor structure fit the data best

\[ \chi^2(26) = 86.44, p < .01, CFI = .95, RMSEA = .08 \]
Perhaps people don't perceive a difference between purpose and process
Can we establish concurrent validity?
Approved for public release; distribution is unlimited.
Both subscales positively correlated with trust

$r = 0.26$

$r = 0.70$
Study 2
Replicate the 2-factor model of understanding & performance
Participants reported experience with digital assistants

“We would like to know about your experience with voice activated digital assistants such as iPhone’s Siri, Microsoft’s Cortana, Amazon’s Alexa, or Google Assistant (“OK Google”). These systems respond to voice commands and execute tasks on your phone or smart home (e.g., Amazon Echo or Google Home). When we talk about digital assistants, we are talking about any of these systems or ones like them.”
Participants completed the TOAST & validation questions

288

Experience
- Ownership
- Knowledge
- Use

TOAST

Trust
Question

Reliance
Question

“I trust the digital assistant”

“I rely on the digital assistant”
Did we replicate the 2-factor model?
Yes! A 2-factor structure fits the data best

\[ \chi^2(26) = 70.90, p < .01, CFI = .98, RMSEA = .05 \]
Approved for public release; distribution is unlimited.
Can we establish concurrent validity?
Yes! Both subscales positively correlated with trust

$r = 0.56$

$r = 0.72$
We demonstrated initial evidence for scale validity

Independent samples

Different systems

Currently collecting OT data
<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I understand what the system should do.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>2.</td>
<td>The system helps me achieve my goals.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>3.</td>
<td>I understand the limitations of the system.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>4.</td>
<td>I understand the capabilities of the system.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>5.</td>
<td>The system performs consistently.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>6.</td>
<td>The system performs the way it should.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>7.</td>
<td>I feel comfortable relying on the information provided by the system.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>8.</td>
<td>I understand how the system executes tasks.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
<tr>
<td>9.</td>
<td>I am rarely surprised by how the system responds.</td>
<td>1 2 3 4 5 6 7</td>
</tr>
</tbody>
</table>