



INSTITUTE FOR DEFENSE ANALYSES

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



The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under Central Research Project C9082, Statistics Working Group. 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

Review was conducted by Rebecca M. Dickinson from the Operational Evaluation Division.

For more information:

Dean Thomas, Project Leader
dthomas@ida.org • (703) 845-6986

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

Copyright Notice

© 2018 Institute for Defense Analyses
4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252.227-7013 (a)(16) [Jun 2013].

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.

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

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.



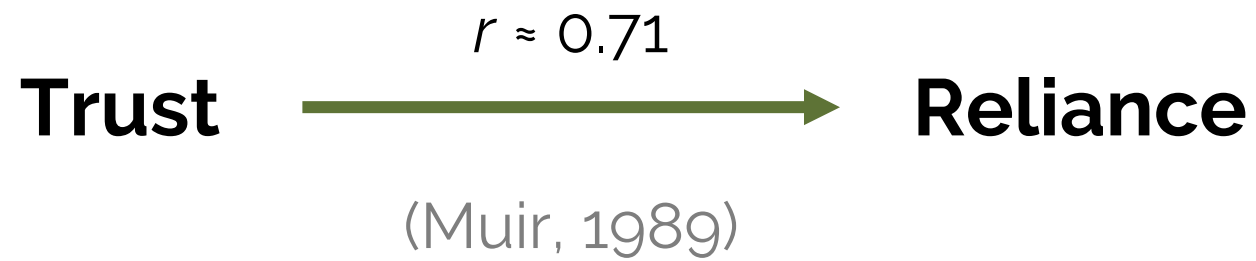
Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

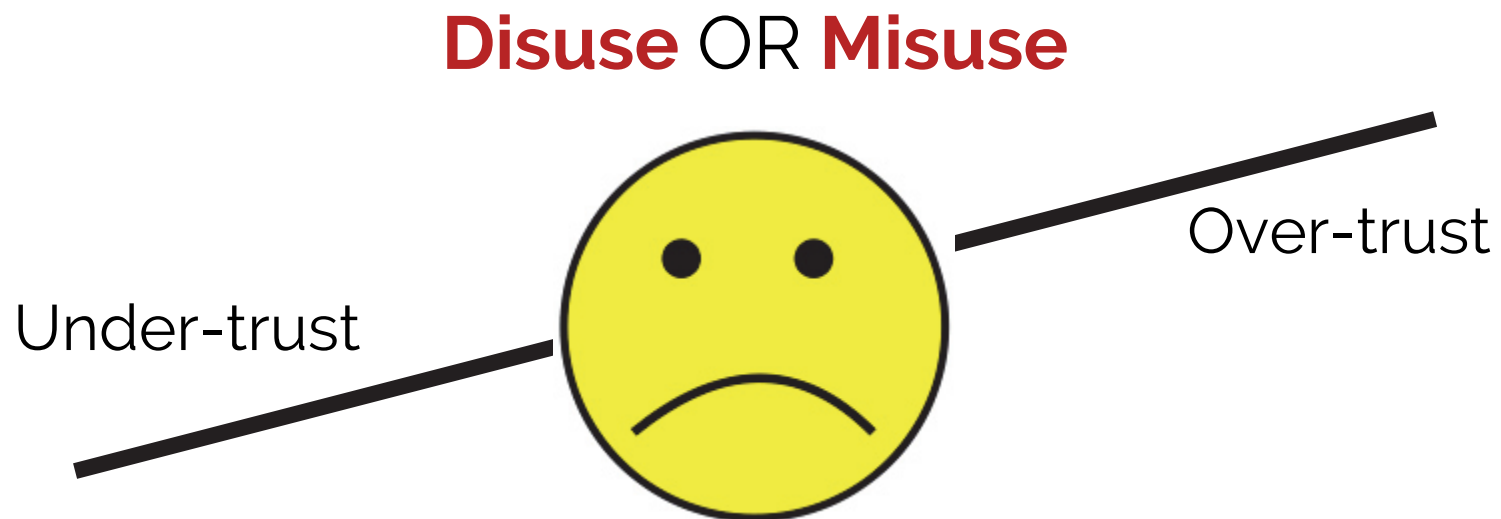
Trust is the expectation that a person or system will help an operator achieve his goal in vulnerable situations

Missile Warning Systems





Mistrust Puts the Operator and the Mission at Risk



Goal: validate a scale to measure
trust in automated systems

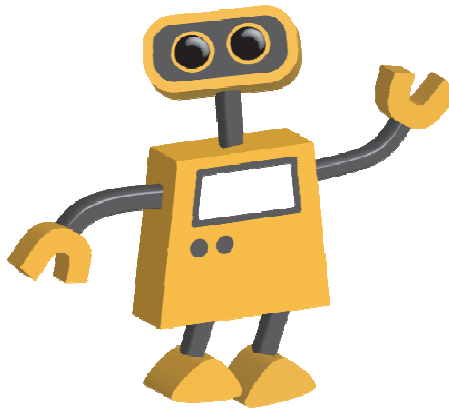
Theory points to 3 factors driving trust in automation

Purpose → Performance → Processes

What you're told
Designer's Intent



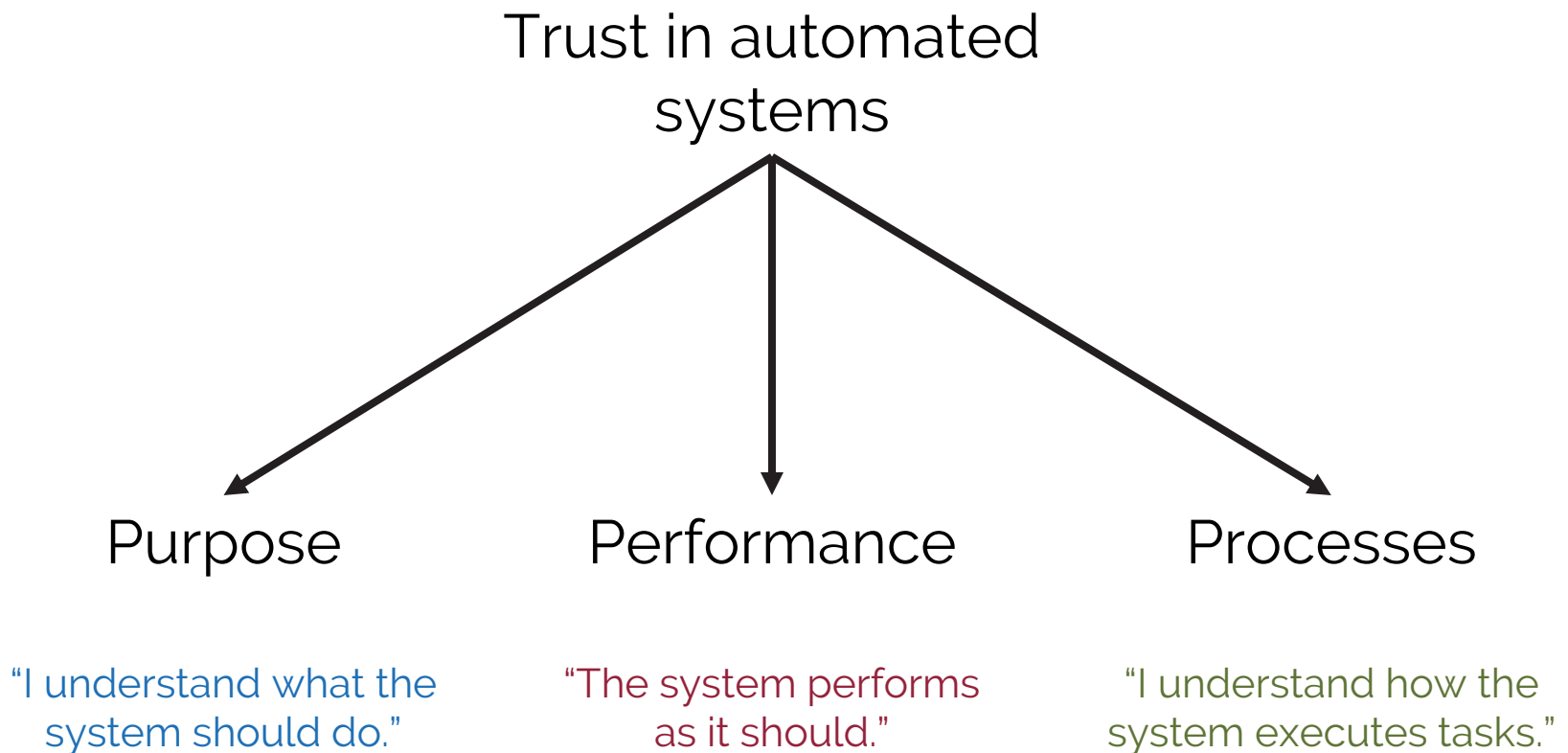
Your experience
Others' experiences



Your expertise



IDA designed a scale to represent these factors



Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

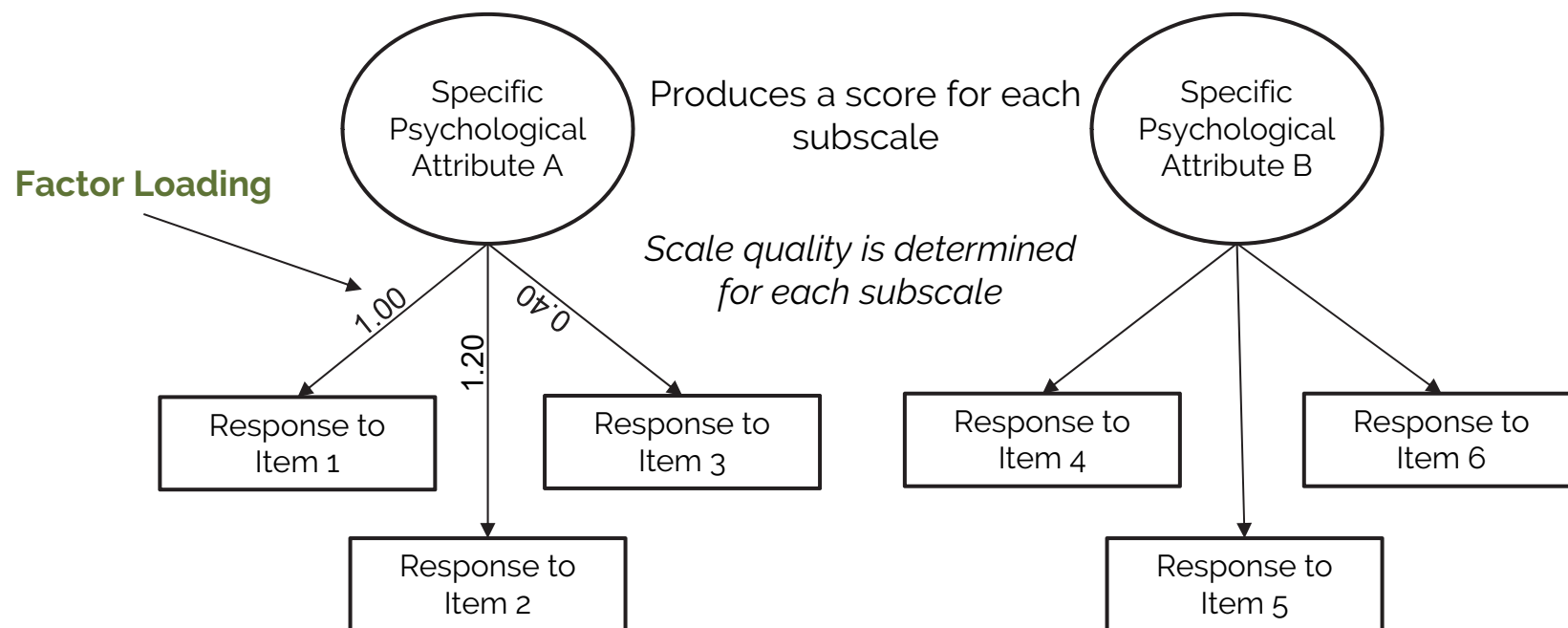
Establishing validity

Approved for public release; distribution is unlimited.

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

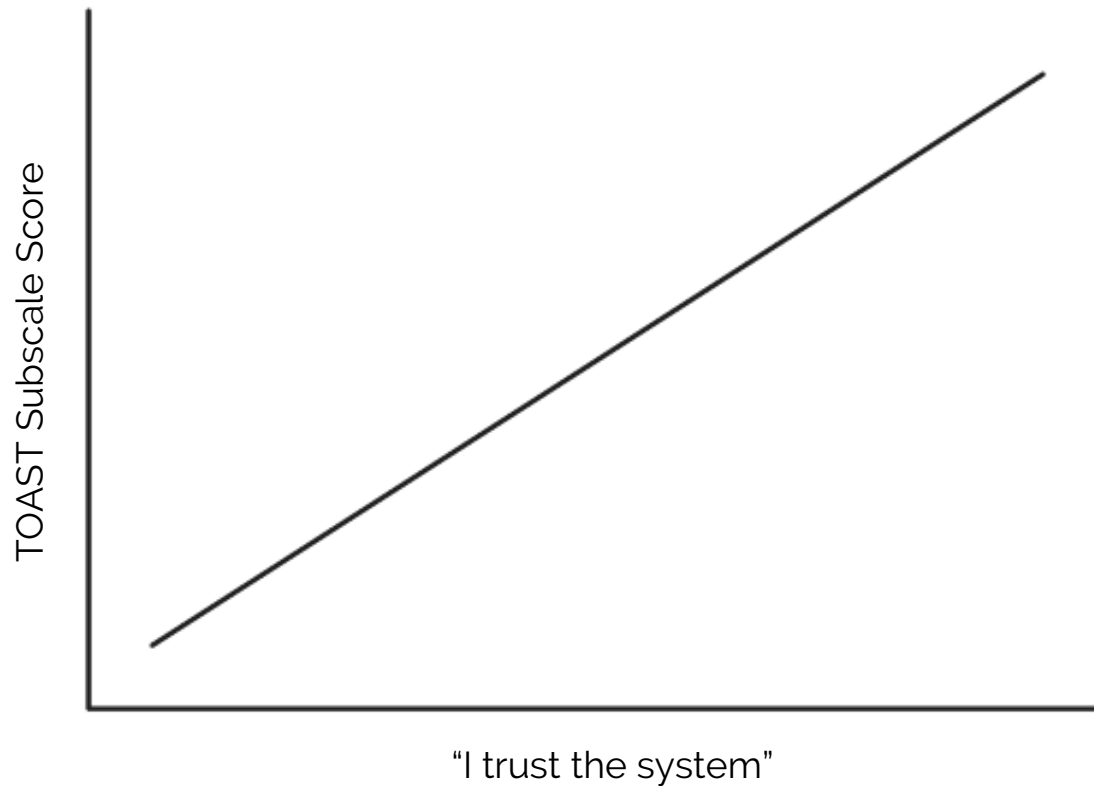


Goodness of fit Indices: χ^2 , *CFI*, and *RMSEA*

We expected items to cluster into 3 factors:
purpose, performance, and process

Establish that the subscales are correlated with theoretically related concepts

Concurrent Validity



We expected subscales to positively correlate with the statement "I trust the system"

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Study 1

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Cadets were randomly assigned to read vignettes



350
Cadets

→ UGV S350
Vignette

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.

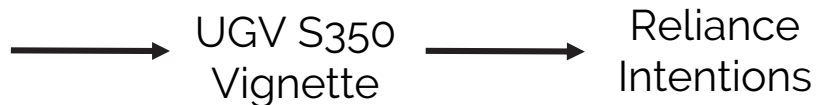
Low Reliability

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



350
Cadets

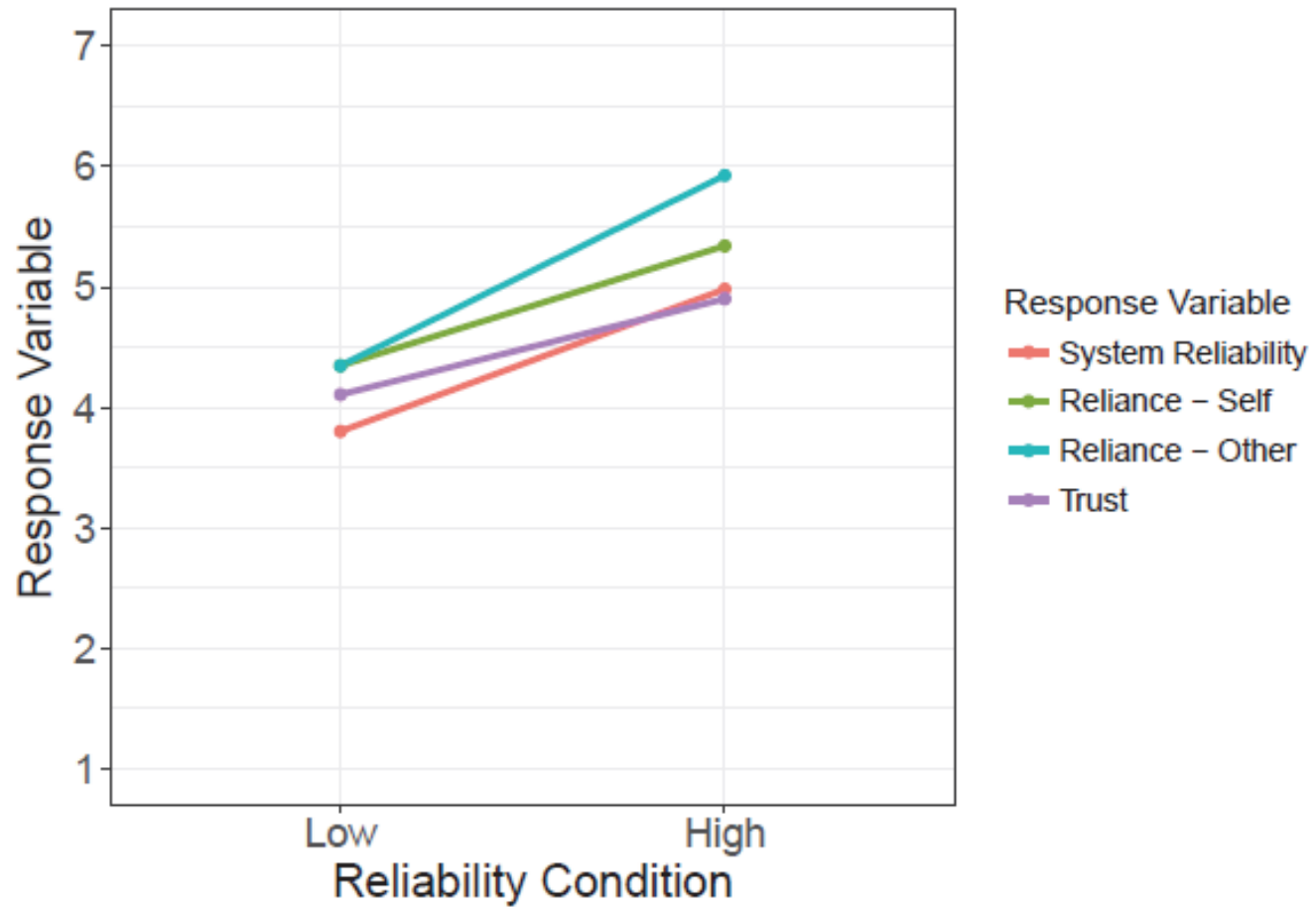


"If placed in a similar situation in real life, I would rely on the UGV S350 to locate victims during search and rescue missions."

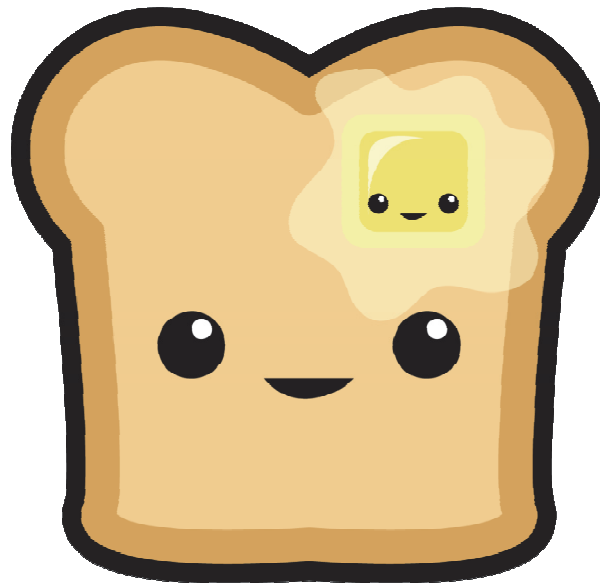
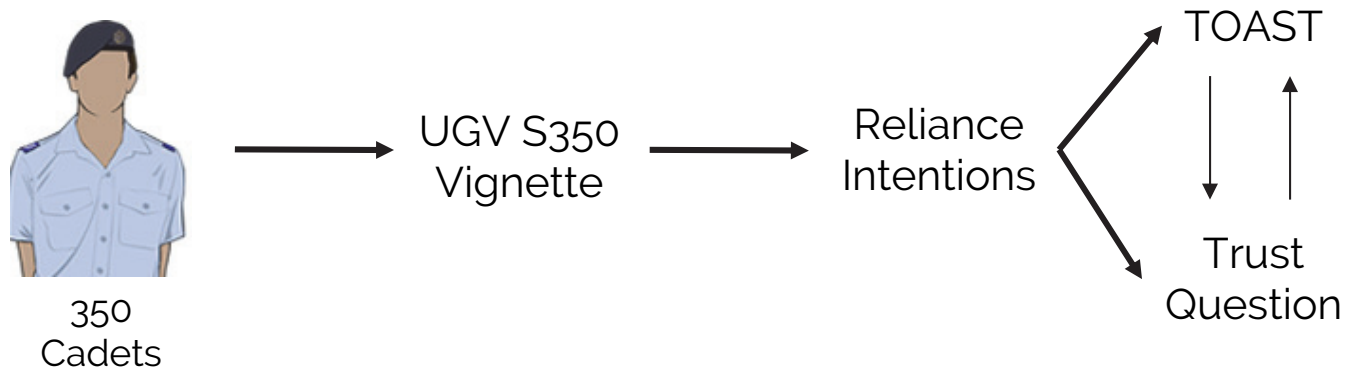
"How reliable is the system?"



Manipulation Check Successful!



Cadets completed the TOAST & validation question



"I trust the system."

How many factors are there?

Approved for public release; distribution is unlimited.

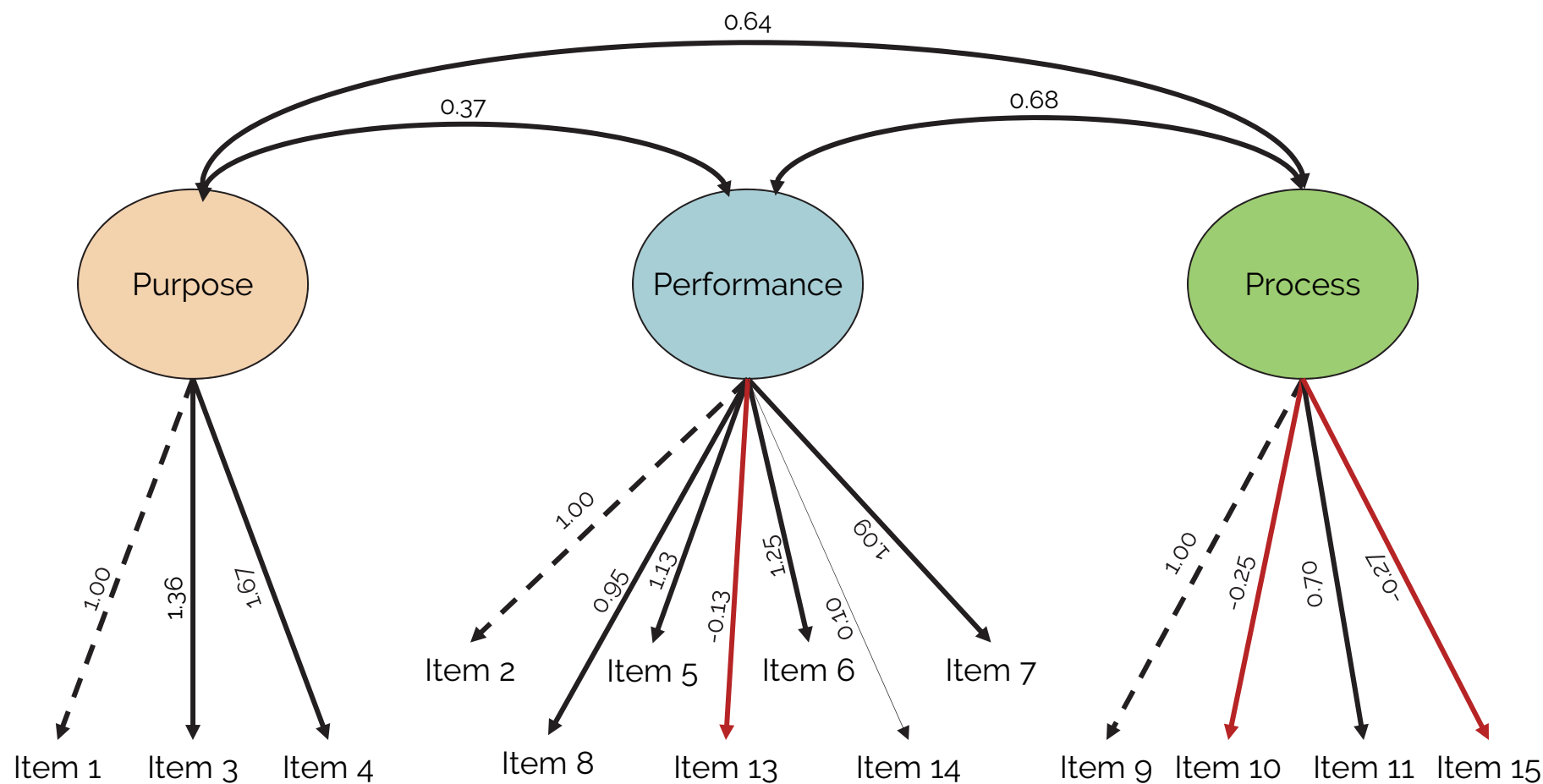
Approved for public release; distribution is unlimited.

We fit a 1- , 2-, and 3-factor model to the data



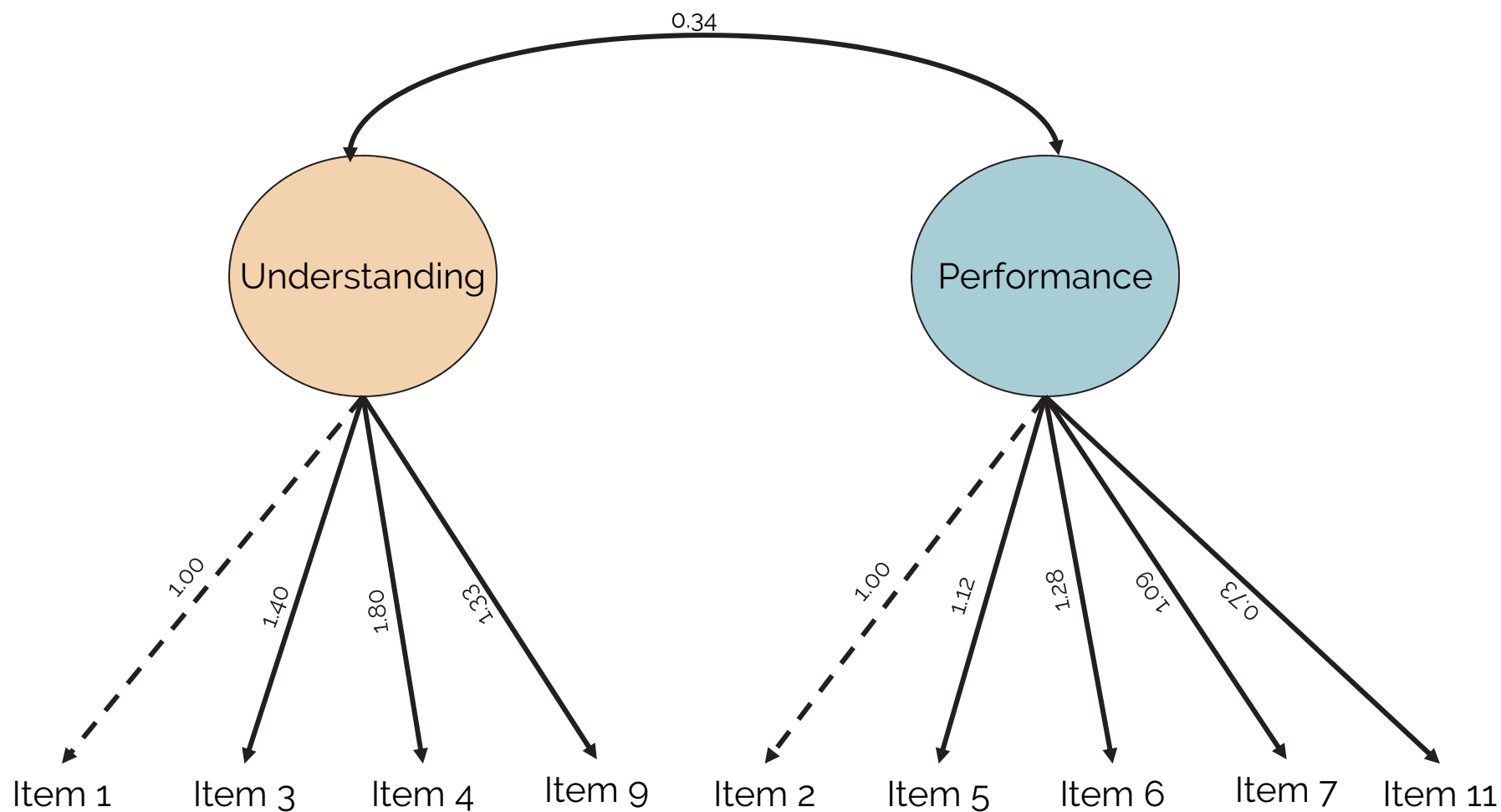
Purpose, Performance, Process

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



$$\chi^2(26) = 86.44, p < .01, CFI = .95, RMSEA = .08$$

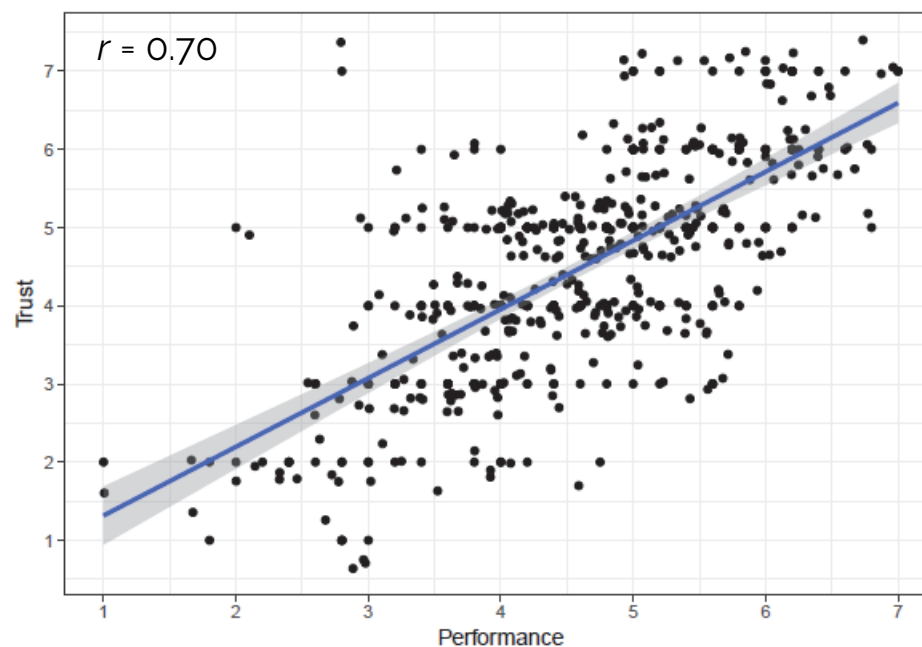
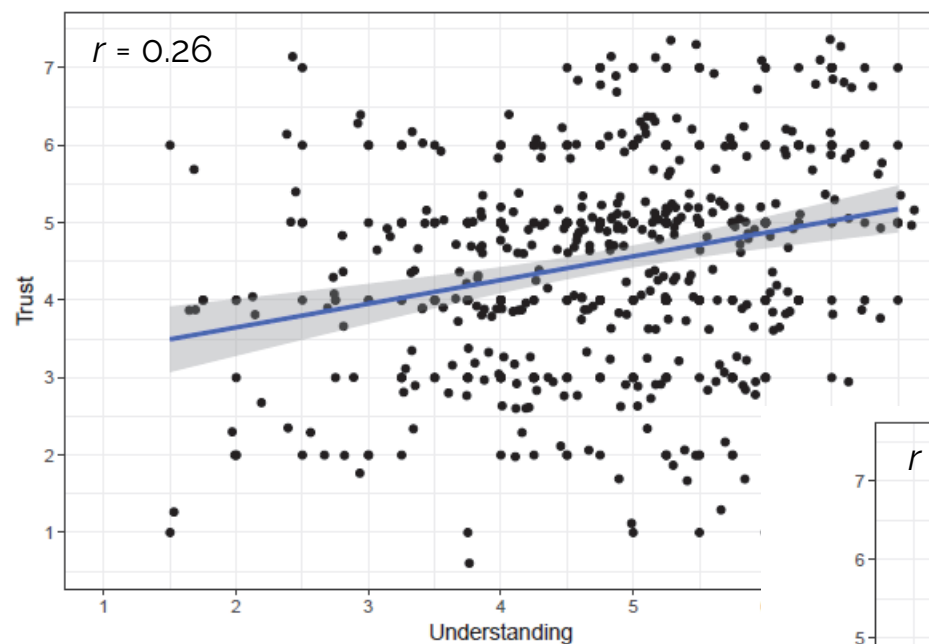
Perhaps people don't perceive a difference between
purpose and process

Can we establish concurrent validity?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Both subscales positively correlated with trust



Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Study 2

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Replicate the 2-factor model of understanding &
performance

Participants reported experience with digital assistants



288



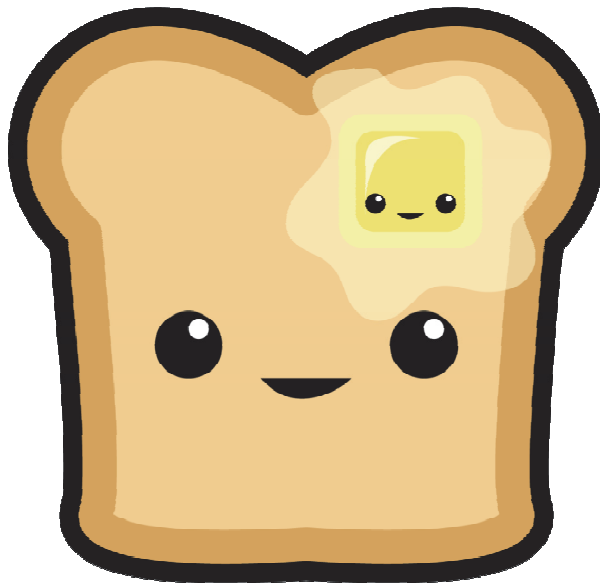
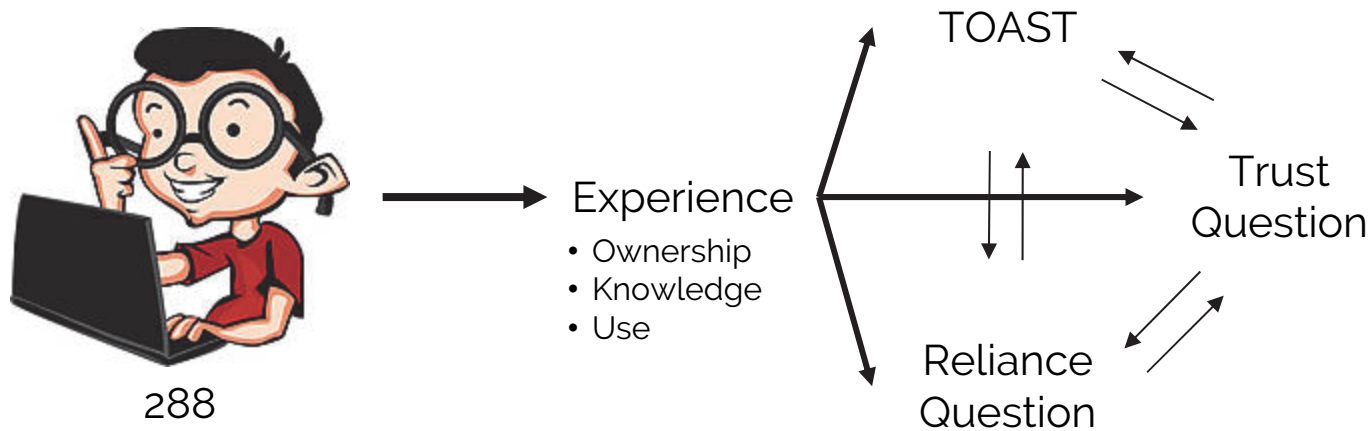
Experience

- Ownership
- Knowledge
- Use



"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



"I trust the digital assistant"

"I rely on the digital assistant"

Approved for public release; distribution is unlimited.

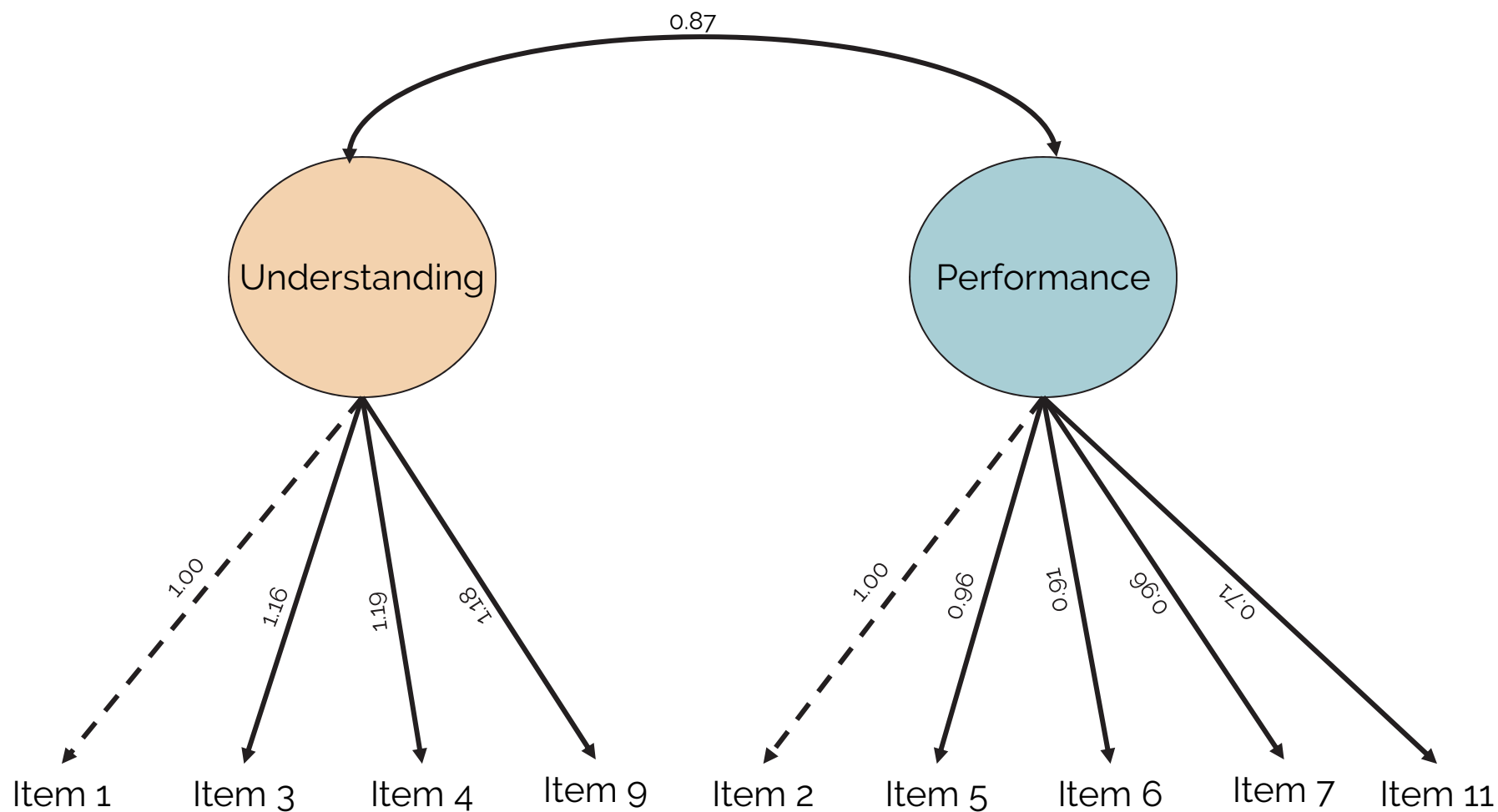
Approved for public release; distribution is unlimited.

Did we replicate the 2-factor model?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

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.

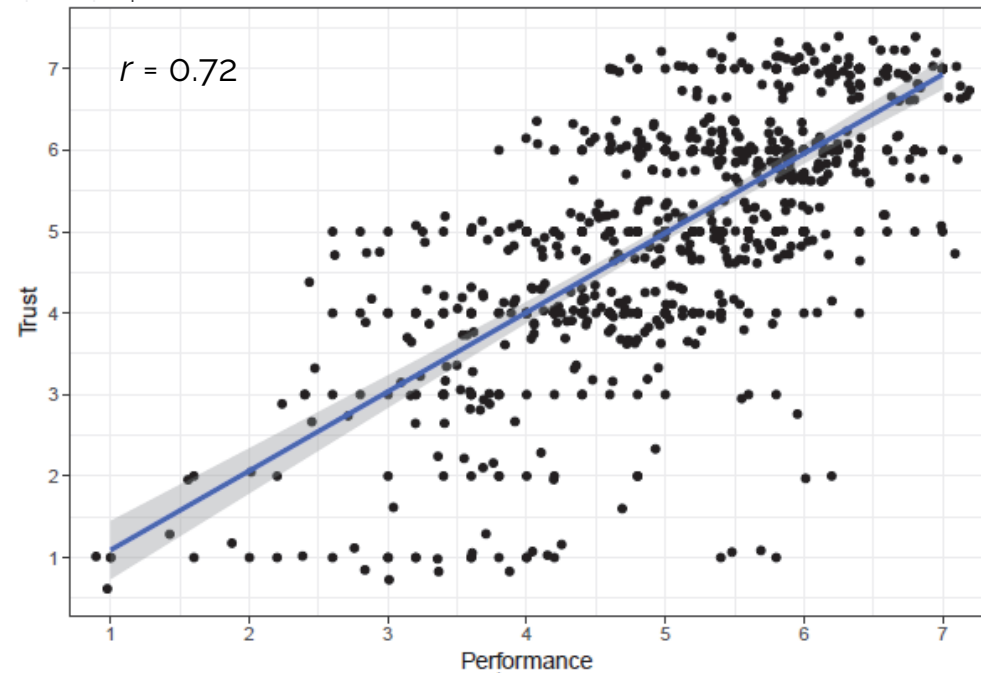
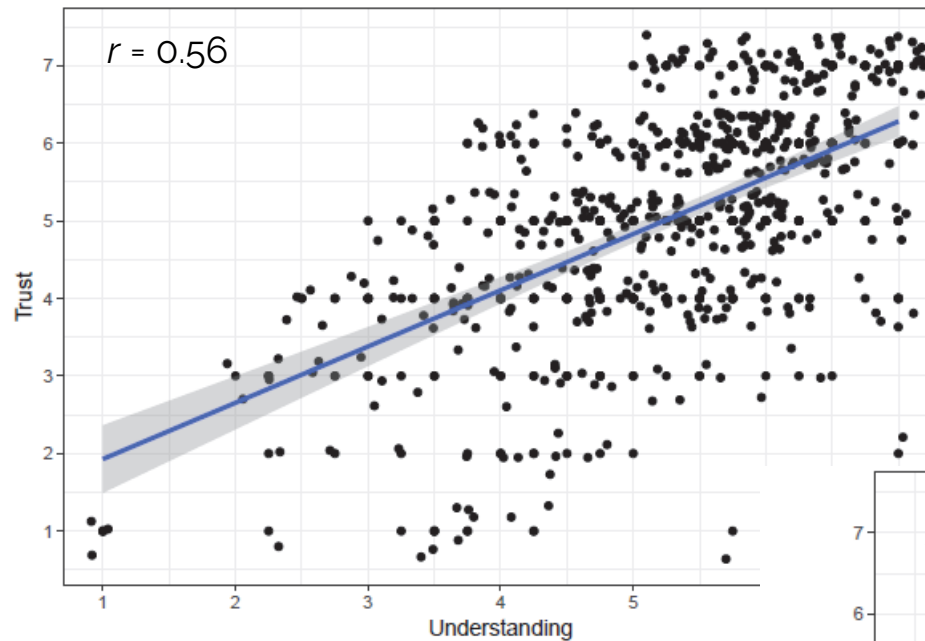
Approved for public release; distribution is unlimited.

Can we establish concurrent validity?

Approved for public release; distribution is unlimited.

Approved for public release; distribution is unlimited.

Yes! Both subscales positively correlated with trust



We demonstrated initial evidence for scale validity

Independent samples

Different systems

Currently collecting OT data

		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