Human performance and embedded intelligent technology in safety-critical systems

Author: Grabowski and Sanborn, 2003
In: International Journal of Human-Computer Studies
58, pp. 637-670
Outline

►► Goal of the Study
►► Background/Definitions
►► Purpose of Study: Marine Transportation
►► Hypotheses and Measures
►► Procedures
►► Data Analysis and Results
►► Study limitations
►► Conclusions
Goal of the Study

To demonstrate, empirically, the impact of embedded intelligent technology on human performance in safety-critical systems.
Embedded Intelligent Systems

These are systems that reside in a larger groups of systems (host).

Therefore, the requirements and performance of the EIS are constrained by the host.

They provide advice, recommendations, conclusions, or explanations in real time

- To the operators (e.g., ship’s masters, mate on watch, ship’s pilot)
- To the hosts
Safety Critical Systems

Examples are:
- Cockpit Automation
- Air Traffic Control
- Space Shuttle Operations
- Nuclear Power Plants
- Medical Systems

Systems that have safety critical implications if
- not functioning properly,
- or not used properly
Shipboard Piloting Expert System (SPES)

► Developed in response to the 1989 Exxon Valdez oil spill by RPI

- The ship ran aground on Bligh Reef because it failed to turn back into the shipping lanes.
- NTSB investigations had cited several human errors among the crew.
Shipboard Piloting Expert System (SPES)

- Provides navigational and pilotage assistance capability
  - would instantly provide warnings to a ship master or pilot of pending hazards
  - and recommend changes in vessel heading to circumvent a hazard
- An embedded intelligent system within a host
  - Host: The ship’s navigation system.
Hypotheses

► Goal: To demonstrate, empirically, the impact of embedded intelligent technology on human performance in safety-critical systems.

► Six Hypotheses
  - Operators will perform better if SPES is available
  - What measures will you look at?
Human Performance measures

- Hypothesis 1 and 2: Decision performance
  - Track keeping – How many cross-track Errors
  - Threat avoidance – How accurately one can identify CPAs
  - Threat maneuvering – Fewer commands communications
    - How much communications/directions are needed
Human Performance measures

- **Hypothesis 3**: Number of alternatives considered
  - Able to make better decisions

- **Hypothesis 4**: Navigational workload
  - How much workload is on the operator (stress)

- **Hypothesis 5**: User decision confidence, decision satisfaction and technology familiarity
  - Perceived Performance

- **Hypothesis 6**: System usage
  - How often they will use the technology
What they did not measure

► Did not measure operators acceptance of the system in the context of:
  - Ease of System Use
    - Tolerance of Nuisances/False Warnings
  - Ease of Learning
    - Ability to retain instructions
  - Perceived Value
    - Do you feel safer with this system?
  - Advocacy of the system
    - Trust
    - Willingness to endorse
Procedure - Subjects

- In a two year period, used 91 subjects that were members of the Exxon Benicia’s bridge watch teams
  1. Ship’s master
  2. Ship’s mate on watch
  3. Ship’s pilot

- All subjects
  - trained on host system
  - used host for 1 year before SPES installation
  - been aboard Exxon tanker an average of 12 years

- Exxon Benicia operates between Valdez, AK and oil terminals on the West Coast of US.
Experimental Design
2x3x3 design

A. Technology Treatment

A1. Host System Alone
A2. Host plus SPES

B. Subjects

B1. Ship’s Master
B2. Mate on Watch
B3. Pilot

C. Voyage Stress

C1. Low
C2. Medium
C3. High
Experimental Design

- Low stress: Clear visibility, no traffic, no wind, no ice, no equipment problems
- Medium stress: An occurrence of traffic, wind or ice
- High stress: Multiple occurrences of bad traffic, weather, or equipment problems

- 16 voyages
  - With two-12 hour transits
  - Each transit was subdivided into 8 subtransit legs
  - Total of 256 possible transit legs
Problems encountered

- Could not get all 256 possible transit legs
  - Due to Heavy Weather
- Could not get equal number of data points in each cell
  - Less voyages during medium and high stress levels.
  - Only 6 of the 91 subjects were exposed to High stress levels.
- Members of the bridge watch team changed over time
  - Researchers said the effects were minimized because of homogeneity of subject pool.
Data Analysis

► Did not use ANOVA tests.

► Used Hotelling’s $T^2$ test
  - Harold Hotelling 1895-1973
  - Wrote a paper on the Student’s t distribution for hypothesis testing (now known as “confidence intervals”).
  - Showed how to use it for looking at multiple variables that are jointly distributed
  - Univariate distribution that is proportional to the F-distribution
Results

- Operators used Embedded Intelligent Technology (EIT) more during medium and high stress situations.
- Operators using EIT in low and medium stress improved on threat avoidance (able to detect the closest point of approach better).
- Operator performance in high stress conditions did not improve with EIT.
Results

- Mates reported higher workload with EIT
- Pilots and Mates required more effort than masters in using EIT.
- Pilots were more confident and satisfied with the host, rather than the EIT.
Limitations of Study

- Could not gather data on all transit legs
- Very few data on technology use and operator performance under high stress
- Statistically supportable conclusions cannot be made.
- Experiment was not random
- Dependencies among observations
- Would have been interesting to see the effect of weather (yes/no), traffic (high, medium, low), on the operators.
Conclusions

- Operators appeared to use the EIT more
  - Novelty effect?
  - Learning effect?
- However, it did not enhance performance in high-stress conditions
- Did improved operator performance in low-stress conditions
  - Could this be because enough data to detect a significant effect?
- Differences were observed in operator’s workload
  - Based on familiarity, confidence, and satisfaction with EIT