

## CLUE

Cross Level User Evaluation

|                                 |   |                        |               |
|---------------------------------|---|------------------------|---------------|
| <b>Programm / Ausschreibung</b> | TAKE OFF, TAKE OFF, TAKEOFF Ausschreibung 2021      | <b>Status</b>          | abgeschlossen |
| <b>Projektstart</b>             | 01.01.2023  | <b>Projektende</b>     | 31.03.2024    |
| <b>Zeitraum</b>                 | 2023 - 2024   | <b>Projektlaufzeit</b> | 15 Monate     |
| <b>Keywords</b>                 | user evaluation; user study; air traffic management |                        |               |

### Projektbeschreibung

Heutzutage stützen sich Evaluierungen neuer Flugsteuerungssysteme in der Regel auf klassische Methoden des Usability-Tests, die aus Beobachtungen in Kombination mit Interviews bestehen. Darüber hinaus gibt es zumindest teilweise automatisierte Systemtests, die sich jedoch eher auf die Arbeitsbelastung und die Stabilität des Systems konzentrieren. Neben der technischen Bewertung stehen heute Arbeitsbelastung, menschliches Fehlverhalten, Teamarbeit und Situationsbewusstsein im Vordergrund. Je nach Art der Tests wird eine große Vielfalt an Methoden gewählt, um Aussagen zu den genannten Themen treffen zu können.

Derzeit wird von den ProbandInnen in Evaluationen ein hoher Grad an Spezialisierung verlangt, um Aussagen über ein System treffen zu können. Insbesondere die hohen Kosten, der organisatorische Aufwand und die begrenzte zeitliche Verfügbarkeit von hochspezialisierten ExpertInnen (FluglotsInnen etc.) stehen einer schnellen, agilen Entwicklung neuer Systeme entgegen. Das daraus resultierende fehlende Feedback durch Nutzerevaluationen und -rückmeldungen stellt ebenfalls ein hohes Entwicklungsrisiko dar: Unzureichendes Feedback erhöht die Gefahr einer verminderten Akzeptanz der entwickelten Software, und es erhöht die Gefahr, Software zu entwickeln, die von den NutzerInnen und dem Markt nicht angenommen wird.

Eine zu untersuchende Alternative sind statistische Methoden, die ursprünglich entwickelt wurden, um Repräsentativitätsmängel durch sogenannte „Hilfsinformationen“ zu korrigieren. Bei diesem alternativen Ansatz wird die Repräsentativität bewusst verletzt, indem aus den oben genannten Gründen (Zeit- und Kostenersparnis) die relevante Zielgruppe durch eine allgemeine Zielgruppe (z.B. freiwillige Studenten) ersetzt wird. Um diesen kostensparenden Fehler zu korrigieren, werden „Hilfsinformationen“ verwendet, um die Unterschiede der Probanden zu charakterisieren und ihre Merkmale aufeinander abzubilden.

Die Lösung dieses Ansatzes erlaubt die Umwandlung von Testergebnissen von Normalpersonen in Testergebnisse von ExpertInnen. Weitere NutzerInnentests können dann einfacher, schneller und kostengünstiger durchgeführt werden. Nur an strategischen Meilensteinen und bei der abschließenden Bewertung können und sollen Tests mit ExpertInnen die „umgewandelten“ Testergebnisse validieren.

### Abstract

Nowadays, evaluations of new flight control systems usually rely on classical methods of usability testing consisting of

observations in combination with interviews. In addition, there are at least partially automated system tests, but these focus more on the workload and stability of the system. In addition to technical evaluation, the focus today is on workload, human error, teamwork, and situational awareness. Depending on the type of tests, a wide variety of methods is selected to be able to make statements on the topics listed.

Currently, a high degree of specialization is required from the test persons in evaluations in order to be able to make statements about a system. In particular, the high costs, organizational effort, and limited time availability of highly specialized experts (air traffic controllers, etc.) stand in the way of rapid, agile development of new systems. The resulting lack of feedback through user evaluations and responses also represents a high development risk: insufficient user feedback increases the danger of reduced acceptance of the developed software, and it increases the danger of developing software that is not accepted by the users and the market.

An alternative to be investigated are statistical methods, which were originally developed to correct deficiencies in representativeness by so-called "auxiliary information". This alternative approach violates representativeness deliberately by replacing the relevant target group with a general target group (e.g., volunteering students) for the reasons mentioned above (time and cost reduction). To correct this cost-saving error, "auxiliary information" is used to characterize the differences of the subjects and to map their characteristics onto each other.

The solution of this approach allows the conversion of test results of ordinary persons to test results of experts. Further user tests can then be performed more easily, faster and more economically. Only at strategic milestones and for the final evaluation, tests with experts can and should validate the "converted" test results.

## **Endberichtkurzfassung**

Die Entwicklung von Flugverkehrskontrollsystemen ist eine Gratwanderung zwischen modernsten Funktionen und unnachgiebigen Sicherheitsanforderungen. Um diese Anforderungen zu erfüllen, sind umfassende Tests unerlässlich. Im Mittelpunkt dieser Testlandschaft steht die Beteiligung der Fluglots:innen selbst, deren Fachwissen sicherstellt, dass die Software den betrieblichen Gegebenheiten und den Bedürfnissen der Endnutzer:innen entspricht. Umfangreiche Benutzertests werden aber durch die Knappheit und Kosten von Fluglots:innen behindert.

In der Sondierung CLUE wurde ein neuartiger Ansatz erforscht, der diese Einschränkung überwinden soll indem Nicht-Expert:innen in die Bewertung von Flugverkehrskontrollsystemen einbezogen werden. Mit Hilfe eines Transformationsmodells, in das Zusatzinformationen aus einem neu entwickelten psychologischen Fragebogen einfließen, wird das Potenzial des Einsatzes von Student:innen zur Vorhersage der Leistung von Fluglots:innen mit einem Flugverkehrskontrollprototyp untersucht. Zwei Methoden, die standardmäßige multiple lineare Regression und die quadratische multiple lineare Regression, wurden für die Erstellung des Transformationsmodells getestet, wobei beide Methoden eine hohe Vorhersagegenauigkeit für die Mehrheit der definierten Leistungsmaße für die Fluglots:innen unter Verwendung der Testergebnisse der Nicht-Expert:innen ermöglichten.

Fragebogenbasierte Messgrößen zeigten jedoch eine geringere Vorhersagegenauigkeit im Vergleich zu Leistungsmessgrößen.

Diese Sondierung unterstreicht die Machbarkeit des Einsatzes von Nicht-Expert:innen für das Testen von Expertensoftware und daraus die Überwindung von Testherausforderungen und die Unterstützung der nutzerzentrierten Designprinzipien.

## **Projektkoordinator**

- Fraunhofer Austria Research GmbH

## **Projektpartner**

- Hochschule für Angewandte Wissenschaften St. Pölten Forschungs GmbH
- 4D Aerospace Research and Simulation GmbH