

# Human Factors – rok 2021, roč. 63

## Číslo 6 (September)



### SPECIAL SECTION: THE SCIENCE BEHIND USABILITY AND UX

**Katie J. Parnell, Victoria A. Banks, Katherine L. Plant, Thomas G. C. Griffin, Peter Beecroft, Neville A. Stanton. *Predicting Design-Induced Error on the Flight Deck: An Aircraft Engine Oil Leak Scenario*. pp. 938–955.**

**Objective:** To explore the types of errors that commercial pilots may make when trying to resolve a suspected engine oil leak using the interfaces currently available. **Background:** The decisions that pilots make often have to be made quickly and under time pressure, with the emphasis on avoiding critical situations from arising. To make the correct decisions, it is vital that pilots have accurate and up-to-date information available. However, interaction with flight deck interfaces may lead to error if they are not effectively designed. **Method:** A hierarchical task analysis was conducted using evidence from pilot interview data to understand the pilots' typical response to a suspected engine oil leak scenario. This was used as the primary input into the Systematic Human Error Reduction and Prediction Approach (SHERPA). **Results:** A total of 108 possible errors were identified. The most common error type was a retrieval error, in which flight crews may retrieve the wrong information about the engine. A number of remedial measures are proposed to try and overcome such issues. **Conclusion:** This analysis provides an initial starting point for identifying potential future design ideas that can assist the pilots in dealing with oil leaks. **Application:** This work has identified the value of applying human error identification methodologies to the assessment of current flight deck processes surrounding engine oil leaks. The method presented permits the operational analysis of possible errors on the flight deck and facilitates the proposition of remedial measures to implement technological innovations that can mitigate error.

- **Keywords:** human systems integration, task analysis, methods and skills, system analysis, system design and analysis (general), human error analysis, accidents, human error, flight displays, aviation and aerospace

**Janine D. Mator, William E. Lehman, Wyatt McManus, Sarah Powers, Lauren Tiller, James R. Unverricht, Jeremiah D. Still. *Usability: Adoption, Measurement, Value*. pp. 956–973.**

**Objective:** We searched for the application of usability in the literature with a focus on adoption, measurements employed, and demonstrated value. Five human factors domains served as a platform for our reflection, which included the last 20 years. **Background:** As usability studies continue to accumulate, there has been only a little past reflection on usability and contributions across a variety of applications. Our research provides a background for general usability, and we target specific usability research subareas within transportation, aging populations, autistic populations, telehealth, and cybersecurity. **Method:** “Usability” research was explored across five different domains within human factors. The goal was not to perform an exhaustive review but, rather, sample usability practices within several specific subareas. We focused on answering three questions: How was usability adopted? How was it measured? How was it framed in terms of value? **Conclusion:** We found that usability is very domain specific. Usability benchmarking studies and empirical standards are rare. The value associated with improving usability ranged widely—from monetary benefits to saving lives. Thus, researchers are motivated to further improve usability practices. A number of data collection and interpretation challenges still call for solutions. **Application:** Findings offer insight into the development of usability, as applied across a variety of subdomains. Our reflection ought to inform future theory development efforts. We are concerned about the lack of established benchmarks, which can help ground data interpretation. Future research should address this gap in the literature. We note that our findings can be used to develop better training materials for future usability researchers.

- **Keywords:** human–computer interaction, interface evaluation, computer interface, product design

**Colton J. Turner, Barbara S. Chaparro, Jibo He. *Typing on a Smartwatch While Mobile: A Comparison of Input Methods*. pp. 974–986.**

**Objective:** The user experience of typing on a smartwatch was evaluated with three unique input methods (tap, trace, and handwriting) while standing and while walking. **Background:** Despite widespread development within the technology industry, smartwatches have had a relatively slow adoption worldwide compared to smartphones. One limiting factor of smartwatches has been the lack of an efficient means of text entry. The 2017 release of Android Wear addressed this issue by providing support for native text entry (i.e., tap, trace, and handwriting). Determining how user performance and subjective ratings compare across these input methods is essential to understanding their contribution to smartwatch user experience. **Method:** Twenty college-age individuals typed phrases using tap, trace, and handwriting input on a smartwatch in three different mobility scenarios (standing, walking a simple course, walking a complex course). **Results:** Participants typed faster with trace (30 words per minute; WPM) than with tap (20 WPM) and handwriting (18 WPM), regardless of mobility. Trace also outperformed tap and handwriting across all subjective metrics, regardless of mobility. **Conclusion:** Trace input appears to be especially well suited for typing on a smartwatch as it was found to be objectively and subjectively superior to tap and handwriting regardless of user mobility. Objectively, typing speeds with trace are shown to be nearly two times faster than most alternative input methods described in the literature. **Application:** Results suggest smartwatch manufacturers should include QWERTY keyboards with trace input as a standard feature in order to provide the best overall typing experience for their users.

- **Keywords:** wearable devices, mobile devices, product design, interface evaluation, usability testing and evaluation

**Philip Kortum, Claudia Ziegler Acemyan, Frederick L. Oswald. *Is It Time to Go Positive? Assessing the Positively Worded System Usability Scale (SUS)*. pp. 987–998.**

**Objective:** The goal of the research presented in this paper was to determine if the positively worded System Usability Scale (SUS) can be used in place of the positively and negatively worded standard SUS instrument for the subjective assessment of usability, and whether the results found here replicate those of Sauro and Lewis. **Background:** Sauro and Lewis' previous study found no evidence that responses to SUS items differed across the standard SUS and the modified, positively worded version of the SUS when participants assessed websites. This study replicates and extends this work by examining a large number of different systems with larger sample sizes to add to the generalizability of previous findings. **Methods:** So that participants could retrospectively assess 20 products, the standard SUS was administered to 268 participants and the positive SUS to 698 participants. SUS scores were computed and the data analyzed using psychometric methods to explore how the two versions of the SUS differed. **Results:** The standard and positive versions of the SUS yielded similar SUS scores. In addition, both versions of the scale demonstrated evidence in support of reliability and validity. **Conclusion:** Either version of the SUS can be used with confidence to measure subjective usability. Furthermore, the scores generated from both versions of the SUS can be directly compared. **Applications:** In situations where cognitive load, participants' spoken language, or item consistency with other surveys being given may be a factor, the positive SUS is a viable alternative to the standard SUS.

- **Keywords:** usability, systems usability scale, SUS, measurement, acquiescence bias

**James R. Lewis. *Measuring User Experience With 3, 5, 7, or 11 Points: Does It Matter?* pp. 999–1011.**

**Objective:** To assess versions of the shorter form variant of Usability Metric for User Experience (UMUX-LITE) questionnaire differing in the number of response options for the items (3, 5, 7, and 11). **Background:** The UMUX-LITE is an efficient (two-item) standardized questionnaire that measures perceived usability. A growing body of evidence shows it closely corresponds to one of the most widely used standardized usability questionnaires, the System Usability Scale (SUS), with regard to both correlation and magnitude of concurrently collected means. Although the "standard" version of the UMUX-LITE uses items with seven response options, there is some variance in practice. **Method:** Members of a corporate user experience panel ( $n = 242$ ) completed surveys rating a recent Web site experience with the SUS and UMUX-LITE, also providing ratings of overall experience and likelihood-to-recommend. **Results:** Scale reliabilities were acceptable (coefficient  $\alpha > .70$ ) with the exception of UMUX-LITE with three response options. All UMUX-LITE correlations with SUS, overall experience, and likelihood-to-recommend were highly significant. For likelihood-to-recommend, there was a significant difference in the magnitude of correlations, with 11 response options higher than three. Although some statistically significant differences were observed in correspondence between SUS and UMUX-LITE scores, these did not seem to translate to practically significant differences. **Conclusion:** The number of UMUX-LITE response options does not matter much, especially in practice. Because the version with three response options showed some weakness with regard to reliability and correlation with likelihood-to-recommend, practitioners should avoid it. **Application:** Unless there is a strong reason to do otherwise, use the "standard" version with seven response options.

- **Keywords:** perceived usability, standardized usability questionnaires, likelihood-to-recommend, UX, response options

**Carlos Silva, Joana Vieira, José C. Campos, Rui Couto, António N. Ribeiro. *Development and Validation of a Descriptive Cognitive Model for Predicting Usability Issues in a Low-Code Development Platform*. pp. 1012–1032.**

**Objective:** The aim of the study was the development and evaluation of a Descriptive Cognitive Model (DCM) for the identification of three types of usability issues in a low-code development platform (LCDP). **Background:** LCDPs raise the level of abstraction of software development by freeing end-users from implementation details. An effective LCDP requires an understanding of how its users conceptualize programming. It is necessary to identify the gap between the LCDP end-users' conceptualization of programming and the actions required by the platform. It is also relevant to evaluate how the conceptualization of the programming tasks varies according to the end-users' skills. **Method:** DCMs are widely used in the description and analysis of the interaction between users and systems. We propose a DCM which we called PRECOG that combines task decomposition methods with knowledge-based descriptions and criticality analysis. This DCM was validated using empirical techniques to provide the best insight regarding the users' interaction performance. Twenty programmers (10 experts, 10 novices) were observed using an LCDP and their interactions were analyzed according to our DCM. **Results:** The DCM correctly identified several problems felt by first-time platform users. The patterns of issues observed were qualitatively different between groups. Experts mainly faced interaction-related problems, while novices faced problems attributable to a lack of programming skills. **Conclusion:** By applying the proposed DCM we were able to predict three types of interaction problems felt by first-time users of the LCDP. **Application:** The method is applicable when it is relevant to identify possible interaction problems, resulting from the users' background knowledge being insufficient to guarantee a successful completion of the task at hand.

- **Keywords:** end-user development, low-code development platforms, descriptive cognitive models, usability, human-computer interaction

## **AUTOMATION, EXPERT SYSTEMS**

**Sarah-Maria Castritius, Christoph Johannes Dietz, Patric Schubert, Johanna Moeller, Simone Morvilius, Sabine Hammer, Chung Anh Tran, Christian T. Haas. *Truck Platooning Under Real Traffic Conditions: First Insights on Behavioral Adaptations and Gap Preference of Professional Drivers*. pp. 1033–1045.**

**Objective:** The aim of the study was to investigate (1) how different gap sizes are perceived by professional truck drivers under real traffic conditions and (2) whether semi-automated platoon driving leads to changes in driving behavior of subsequent manual driving. **Background:** Platoon driving is a current branch in the development of automated driving in which two or more vehicles build a convoy. The lead vehicle is controlled manually while following vehicles are electronically coupled to it and drive semi-automated with small gaps in order to achieve a better traffic flow and potential fuel savings. **Method:** In a real road experiment, 10 trained professional truck drivers completed a total of 33 test drives with a two-truck platoon on the German highway A9 with a gap size of either 15 or 21 m, in the leading and the following vehicle. **Results:** (1) The drivers experienced both gap sizes as comfortable and preferred the smaller gap size of 15 m. (2) Both gap sizes led to significantly higher standard deviation of lane position in post- compared to pre-platoon driving. No significant difference in distance keepings in post- compared to pre-platoon driving occurred. Qualitative data give hints on difficulties, when switching back to regular truck driving. **Conclusion:** The results

implicate that small gap sizes are perceived as comfortable by drivers and that platoon driving has an influence on subsequent manual driving. **Application:** Countermeasures to behavioral adaptations should be considered in order to ensure a safe conduction of platoon driving.

- **Keywords:** truck platoon driving, commercial truck driver, on-road study, human–automation interaction, driver behavior, vehicle automation

## COGNITION

**Brittany N. Neilson, Curtis M. Craig, Raelyn Y. Curiel, Martina I. Klein.** *Restoring Attentional Resources With Nature: A Replication Study of Berto's (2005) Paradigm Including Commentary From Dr. Rita Berto.* pp. 1046–1060.

**Objective:** The aim of this study is to replicate Berto's (2005) heavily cited work on attention restoration. **Background:** Nature interventions have gained increased interest for improving performance of attentionally demanding tasks. Berto (2005) indicated that viewing digital nature images could improve performance on a subsequent response inhibition task, the sustained attention to response task (SART). However, experimental design and statistical concerns about her experiments as well as failure to support her findings across multiple unpublished studies in our laboratory provided rationale for this replication study. **Method:** Twenty participants were each assigned to one of three digital image conditions: nature, urban, and control. Participants performed the SART before and after digital image exposure. **Results:** SART performance metrics (total correct target responses, mean response time, and transformed  $d'$ ) were analyzed using 2 (SART)  $\times$  3 (image interventions) mixed design ANOVAs. The results failed to replicate Berto (2005). **Conclusion:** Possible reasons for not replicating Berto (2005) are discussed, including (1) sample differences, (2) different testing environments and procedures, (3) insufficient attentional depletion, and (4) individual differences. **Applications:** Research needs to determine the effectiveness of such interventions, the specific attention tasks that might benefit, and the individual difference variables relevant for attention restoration.

- **Keywords:** attentional processes, environmental design, fatigue, multiple resource models, replication

## HUMAN-ROBOT INTERACTION

**Tjaša Kermavnar, Kevin J. O'Sullivan, Adam de Eyto, Leonard W. O'Sullivan.** *Relationship Between Interface Pressures and Pneumatic Cuff Inflation Pressure at Different Assessment Sites of the Lower Limb to Aid Soft Exoskeleton Design.* pp. 1061–1075.

**Objective:** The aim was to develop a means of predicting interface pressure from cuff inflation pressure during circumferential compression at the lower limb, in order to inform the design of soft exoskeletons. **Background:** Excessive mechanical loading of tissues can cause discomfort and soft tissue injury. Most ergonomic studies on exoskeletons are of interface pressure, but soft exoskeletons apply circumferential pressures similar to tourniquet cuffs by way of cuff inflation pressure. This study details the relationship between interface and cuff inflation pressures for pneumatic tourniquet cuffs. **Method:** Pneumatic cuffs of different widths were inflated to target pressures on (A) a rigid

cylinder, (B) the dominant thigh and calf, and (C) knee of healthy participants standing still. Interface pressures were measured under the cuffs using a pressure-sensing mat. Average interface pressures were then compared to cuff inflation pressures. The influence of cuff width, cuff inflation pressure, and participants' anthropometric data on pressure transmission was assessed. **Results:** A strong linear relationship between cuff inflation pressures and interface pressures was observed. Interface pressures were generally higher than cuff inflation pressures. The efficiency of pressure transmission to the lower limb depended on assessment site, adipose tissue thickness, cuff size, cuff inflation pressure, and possibly limb circumference. Regression equations were developed to predict interface pressures at the thigh, calf, and knee. **Conclusion:** Interface pressures under pneumatic cuffs are influenced by the cuff size, cuff inflation pressure, and tissue compressibility. Predicted interface pressure from cuff inflation pressure and vice versa can be used to aid the design of soft exoskeletons.

- **Keywords:** soft exoskeleton–human contact, interface pressure, cuff inflation pressure, pressure transmission efficiency

## INDIVIDUAL DIFFERENCES

**Johannes Kraus, David Scholz, Martin Baumann. *What's Driving Me? Exploration and Validation of a Hierarchical Personality Model for Trust in Automated Driving.* pp. 1076–1105.**

**Objective:** This paper presents a comprehensive investigation of personality traits related to trust in automated vehicles. A hierarchical personality model based on Mowen's (2000) 3M model is explored in a first and replicated in a second study. **Background:** Trust in automation is established in a complex psychological process involving user-, system- and situation-related variables. In this process, personality traits have been viewed as an important source of variance. **Method:** Dispositional variables on three levels were included in an exploratory, hierarchical personality model (full model) of dynamic learned trust in automation, which was refined on the basis of structural equation modeling carried out in Study 1 (final model). Study 2 replicated the final model in an independent sample. **Results:** In both studies, the personality model showed a good fit and explained a large proportion of variance in trust in automation. The combined evidence supports the role of extraversion, neuroticism, and self-esteem at the elemental level; affinity for technology and dispositional interpersonal trust at the situational level; and propensity to trust in automation and a priori acceptability of automated driving at the surface level in the prediction of trust in automation. **Conclusion:** Findings confirm that personality plays a substantial role in trust formation and provide evidence of the involvement of user dispositions not previously investigated in relation to trust in automation: self-esteem, dispositional interpersonal trust, and affinity for technology. **Application:** Implications for personalization of information campaigns, driver training, and user interfaces for trust calibration in automated driving are discussed.

- **Keywords:** trust in automation, human-automation interaction, trust formation, personality, affinity for technology, structural equation modeling

## SURFACE TRANSPORTATION

**Taylor Shupsky, Adriana Lyman, Jibo He, Maryam Zahabi. *Effects of Mobile Computer Terminal Configuration and Level of Driving Control on Police Officers' Performance and Workload.* pp. 1106–1120.**

**Objective:** The objective of this study was to assess police officers' performance and workload in using two mobile computer terminal (MCT) configurations under operational and tactical driving conditions. **Background:** Crash reports have identified in-vehicle distraction to be a major cause of law enforcement vehicle crashes. The MCT has been found to be the most frequently used in-vehicle technology and the main source of police in-vehicle distraction. **Method:** Twenty police officers participated in a driving simulator-based assessment of driving behavior, task completion time, and perceived workload with two MCT configurations under operational and tactical levels of driving. **Results:** The findings revealed that using the MCT configuration with speech-based data entry and head-up display location while driving improved driving performance, decreased task completion time, and reduced police officers' workload as compared to the current MCT configuration used by police departments. Officers had better driving but worse secondary task performance under the operational driving as compared to the tactical driving condition. **Conclusion:** This study provided an empirical support for use of an enhanced MCT configuration in police vehicles to improve police officers' safety and performance. In addition, the findings emphasize the need for more training to improve officers' tactical driving skills and multitasking behavior. **Application** The findings provide guidelines for vehicle manufacturers, MCT developers, and police agencies to improve the design and implementation of MCTs in police vehicles considering input modality and display eccentricity, which are expected to increase officer and civilian safety.

- **Keywords:** in-vehicle technology, law enforcement, distraction, safety, accidents

## **SPECIAL SECTION: MEASURING COGNITIVE WORKLOAD IN HUMAN FACTORS**

**Francesco N. Biondi, David L. Strayer, William J. Horrey, Joel M. Cooper, Joel A. Cort. [Preface to the Special Section on Measuring Cognitive Workload in Human Factors.](#) pp. 1121–1124.**

Cognitive workload is a reflection of the demands of a task and the attention and effort allocated to the task by the operator. Both the task demands and the effort invested by the operator can fluctuate over time. Measuring cognitive load during the human interaction within the workplace is one of the most pressing challenges in human factors and ergonomics (HF/E). The HF/E community has long investigated the effect of nonoptimal cognitive workload on manually operated machines (Brookhuis et al., 1991; Norman & Rumelhart, 1983). However, as systems become more automated and the role of the human shifts toward that of a supervisor and away from that of an operator, the need to accurately monitor cognitive underload and overload—both of which can be detrimental to performance—is paramount for a safe and effective human–system cooperation.