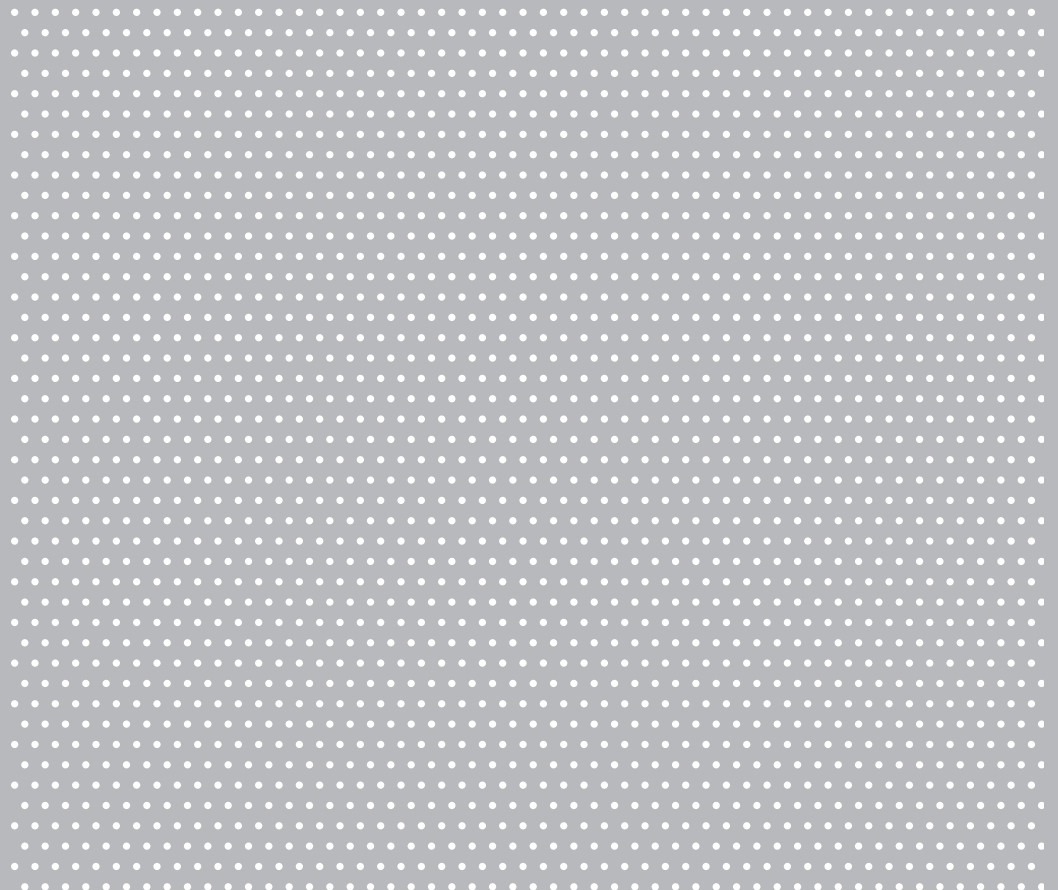


**Beurteilung von Aufgabenlasten von
digitalen Informationssystemen auf
Flurförderzeugen: Datenbrille (HMD)
vs. Monitor (Grundlagenuntersuchung)**



Verfasst von: Benno Gross, Jörg Rissler
Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung e. V.,
Sankt Augustin

Herausgegeben von: Deutsche Gesetzliche Unfallversicherung e. V. (DGUV)
Glinkastr. 40
10117 Berlin
Telefon: +49 30 130010
Telefax: +49 30 130019876
Internet: www.dguv.de
E-Mail: info@dguv.de

– Oktober 2018 –

Layout und Gestaltung: Deutsche Gesetzliche Unfallversicherung e. V. (DGUV)

ISBN: 978-3-86423-221-3
ISSN: 2190-7994

Kurzfassung

Beurteilung von Aufgabenlasten von digitalen Informationssystemen auf Flurförderzeugen: Datenbrille (HMD) vs. Monitor (Grundlagenuntersuchung)

Neue digitale Arbeitsmittel bestimmen zunehmend die Arbeitswelt. Sie ermöglichen oftmals effizientere, mobilere und flexiblere Arbeitsabläufe; gleichzeitig können aber auch neue Gefahren und Risiken für die Beschäftigten entstehen. Aus diesem Grund befasst sich das vorliegende Forschungsprojekt mit den möglichen Auswirkungen des Einsatzes von Datenbrillen und Monitoren auf Gabelstaplern auf die sensorische und kognitive Belastung der Beschäftigten. Hierzu wurde im Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA) eine möglichst realitätsnahe Fahrsimulation aufgebaut.

Diese wird eingesetzt, um in einem Studienkollektiv Reaktionszeiten nach einem standardisierten Testverfahren zu messen (Detection Response Task (DRT) nach ISO 17488). Außerdem wird die subjektiv wahrgenommene Belastung der Probanden mit dem etablierten Fragebogen NASA Task Load Index (NASA-TLX) erfasst. Im Studienkollektiv wurde eine Zunahme der sensorischen und kognitiven Belastung durch das Nutzen von Anzeigegeräten nachgewiesen. Unterschiede zwischen verschiedenen Arten von Anzeigegeräten traten nicht auf. Wie sich diese Belastung auf die Sicherheit und Gesundheit von Beschäftigten auswirkt, müssen weitere Studien klären.

Abstract

Assessment of employee task load due to digital information systems on industrial trucks: head-mounted display (HMD) vs. monitor (basic study)

The working environment is being affected increasingly by new digital devices. Such devices often make work processes more efficient, mobile and flexible, although they are also capable of posing new risks and hazards to employees. For this reason, the present research project is concerned with the possible effects of the use of head-mounted displays and monitors on forklift trucks in terms of sensory and cognitive task load for employees. To this end, a highly realistic driving simulator has been set up at the Institute for Occupational Safety & Health of the German Social Accident Insurance (IFA). This is used for measuring the response times of a study population using a standardised test method (Detection Response Task (DRT) conforming to ISO 17488).

The subjectively perceived stressing of test subjects is also recorded with the established NASA Task Load Index (NASA-TLX). In the study population an increase was ascertained in sensory and cognitive task load due to the use of display devices. Differences between different types of display devices did not arise. What possible effect this task load has on the health and safety of employees remains to be determined in further studies.

Résumé

Évaluation des charges de travail causées par les systèmes d'information numériques sur les chariots de manutention : lunettes connectées (HMD) vs. écran d'affichage (étude fondamentale)

Les nouveaux outils de travail numériques jouent un rôle de plus en plus important dans le monde du travail. Permettant souvent de travailler avec davantage de rapidité, de flexibilité et de mobilité, ils peuvent être en même temps à l'origine de dangers et de risques nouveaux pour les employés. C'est pourquoi le présent projet de recherche est consacré à l'impact que l'utilisation de lunettes connectées et d'écrans d'affichage sur les chariots de manutention est susceptible d'avoir sur la charge sensorielle et cognitive des employés. À cet effet, l'Institut pour la sécurité et la santé au travail de la DGUV (IFA) a réalisé une simulation de conduite aussi proche que possible de la réalité. Celle-ci sert à mesurer les temps de réaction, auprès d'un groupe de sujets, selon une méthode d'essai standardisée (Tâche de Détection-Réponse (DRT) selon ISO 17488).

De plus, la charge perçue subjectivement par les sujets est saisie à l'aide d'un questionnaire couramment utilisé: le NASA Task Load Index (NASA-TLX). Dans le groupe de sujets ayant participé à l'étude, il a été démontré que l'utilisation des appareils d'affichage provoquait une augmentation de la charge sensorielle et cognitive. Aucune différence n'a été constatée entre les différents types d'appareils d'affichage. D'autres études devront se pencher sur la question de savoir quel est l'impact de cette charge sur la sécurité et la santé des employés.

Resumen

Evaluación de las repercusiones de los sistemas de información digitales en las carretillas elevadoras: gafas de datos (HMD) o monitor (investigación de base)

Los nuevos instrumentos de trabajo digitales determinan cada vez más el mundo laboral. Con frecuencia, permiten introducir procesos más eficientes, móviles y flexibles; pero, al mismo tiempo, también pueden albergar nuevos peligros y riesgos para los empleados. Por ese motivo, este proyecto de investigación estudia los posibles efectos del empleo de las gafas de datos y monitores en las carretillas elevadoras para la carga sensorial y cognitiva de los empleados. Al respecto se ha realizado en el instituto de seguridad laboral del seguro alemán de accidentes (Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung, IFA) una simulación de conducción lo más realista posible. Dicha simulación se emplea para medir en un colectivo de estudio los tiempos de reacción en función de un procedimiento de pruebas estandarizado (Detection Response Task, DRT, según ISO 17488).

Asimismo, la carga percibida subjetivamente por los sujetos del estudio se recoge con el cuestionario establecido basado en el NASA Task Load Index (NASA-TLX). En el colectivo de estudio se ha demostrado la existencia de un aumento de la carga sensorial y cognitiva debido al empleo de indicadores electrónicos. No se han detectado diferencias entre los diversos tipos de indicadores. Otros estudios posteriores deberán determinar de qué manera dicha carga puede afectar a la seguridad y la salud de los empleados.

Inhaltsverzeichnis

1	Einleitung.....	11
1.1	Hintergrund der Studie	11
1.2	Hintergrund: Datenbrillen in der Arbeitswelt	11
1.3	Laborstudie von BGHW und IFA: Anzeigesysteme auf Gabelstaplern.....	11
1.4	Forschungsfrage.....	12
2	Methodik	13
2.1	Studienkollektiv – Beschreibung	13
3	Versuchsablauf.....	15
3.1	Messung der Reaktionszeiten.....	15
3.2	Gabelstaplersimulator.....	16
3.3	Displays	16
3.4	Fragebögen	17
4	Ergebnisse	19
4.1	Ergebnisse – Allgemein	19
4.2	Ergebnisse – Fahrfehler.....	19
4.3	Ergebnisse – Reaktionstest.....	20
4.3.1	Ergebnisse – Reaktionstest – Allgemein	20
4.3.2	Ergebnisse – Reaktionstest – Belastungszunahme bei Gerätenutzung	21
4.3.3	Ergebnisse – Reaktionstest – Belastungsunterschiede zwischen unterschiedlichen Geräten	23
4.4	Ergebnisse – NASA-TLX.....	24
4.4.1	Ergebnisse – NASA-TLX – Belastungszunahme bei Gerätenutzung.....	24
4.4.2	Ergebnisse – NASA-TLX – Belastungsunterschiede zwischen unterschiedlichen Geräten	26
4.4.3	Ergebnisse – NASA-TLX – Einzelfragen	27
4.5	Ergebnisse – Masteinblendung.....	28
4.6	Ergebnisse – Allgemeiner Fragebogen	29
5	Diskussion	31
5.1	Studienkollektiv	31
5.2	Fahrfehler.....	31
5.3	Masteinblendung	31
5.4	Reaktionszeiten	31
5.5	NASA-TLX.....	32
5.6	Allgemeine Fragen	32
6	Schlussfolgerung.....	33
7	Literatur	35
	Anhang 1: Probandeninformation und Einverständniserklärung.....	37
	Anhang 2: Fragebögen	39
	Vor den Fahrten	42
	Nach jeder Fahrt	42
	Fragen zum Abschluss.....	42

Danksagung

Dieses Forschungsprojekt wäre nicht möglich gewesen ohne den zum Teil sehr zeitintensiven Einsatz vieler Beschäftigter der Berufsgenossenschaft Handel und Warenlogistik (BGHW), des Instituts für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA) und des Projektpartners CINOVIATION GmbH.

Besonderer Dank gilt *Marieke Dahl* und *Dr. Hans-Peter Kany* (BGHW), *Marcel Rahyr* (CINOVIATION GmbH) sowie allen Mitarbeitern des IFA und von der Fa. CINOVIATION GmbH, die bei dem Forschungsprojekt als Probanden mitgewirkt haben.

Außerdem danken wir *Dr. Michael Bretschneider-Hagemes* und *Andreas Stefan*, die im Rahmen ihrer Tätigkeit im IFA maßgeblich an der Planung, der Durchführung und dem Erfolg des Forschungsprojekts mitgewirkt haben.

1 Einleitung

1.1 Hintergrund der Studie

Bei Kommissionier-Tätigkeiten im Regallager können Beschäftigte auf verschiedene Arten Informationen zu ihrem Arbeitsablauf erhalten. In den letzten Jahren wurden dazu akustische (Pick-by-Voice) oder visuelle (Pick-by-Vision) Unterstützungssysteme entwickelt, wobei die Beschäftigten auch Flurförderzeuge nutzen.

Wer einen Gabelstapler fährt, erhält die für diesen Arbeitsbereich relevanten Informationen meist über ein Monitoranzeigesystem in der Fahrgastzelle. Dabei können die Informationsaufnahme während des Fahrens und die damit verbundenen Blickabwendungen vom Fahrgeschehen zu einem erhöhten Unfallrisiko führen. Durch den Einsatz von Datenbrillen an Gabelstaplerfahrer-Arbeitsplätzen mit Einblendungen von Anweisungen und Informationen im Sichtfeld und der damit verbundenen Reduktion von Blickabwendungen vom Fahrgeschehen erwartet das Unternehmen neben einer effizienteren Arbeitsverrichtung auch eine mögliche Reduktion von ablenkungsbedingten Gefahrenpotenzialen an Arbeitsplätzen auf Gabelstaplern.

Diese Studie untersucht in einem ersten Schritt kognitive Belastungen, die auftreten können, wenn man verschiedene Displays (Tablet, Datenbrille) zur Informationsaufnahme und Interaktion beim Fahren von Gabelstaplern nutzt. Damit kann eine Grundlage gelegt werden, um weitere Auswirkungen auf die Sicherheit und Gesundheit von Beschäftigten zu bestimmen.

Die Projektpartner in diesem Projekt waren die Berufsgenossenschaft Handel und Warenlogistik (BGHW), das Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA) und die Fa. CINOVIATION.

1.2 Hintergrund: Datenbrillen in der Arbeitswelt

Datenbrillen sind kopfgetragene Micro-Computer (Wearables), die über optische Displays im Blickfeld der Person verfügen und somit als Head-Mounted Displays (HMD) klassifiziert werden. Ähnlich wie Smartphones verfügen sie über ein Betriebssystem, in der Regel über Bluetooth- und WLAN-Sensoren, sowie eine Kamera, mit der sich Videos und Fotos aufzeichnen und anzeigen lassen. Bei der Nutzung von Datenbrillen wird die Realität durch virtuelle Einblendungen von Informationen oder Objekten in Echtzeit durch Überlagerung der Wahrnehmungsebenen angereichert. Deshalb spricht man im Zusammenhang mit Datenbrillen, je nach Ausprägung der Wahrnehmungsüberlagerung, von Assisted Reality, Augmented Reality

oder Mixed Reality – also einer unterstützten, erweiterten oder vermischten Realität.

Die Potenziale dieser Technologie stoßen auf großes Interesse in verschiedenen Bereichen der Arbeitswelt. Das US-Marktforschungsunternehmen Gartner geht daher davon aus, dass 20 % der großen Unternehmen bis 2019 mit der Implementierung von Virtual- oder Augmented-Reality-Lösungen beginnen werden [1]. Insbesondere die Intralogistik mit ihren komplexen Arbeitsprozessen, den hohen Anforderungen an Mobilität und Flexibilität kann dabei besonders stark vom technologischen Wandel profitieren. Bereits heute können Kommissionierszenarien durch den Einsatz von Datenbrillen in Form von Pick-by-Vision-Systemen (Assisted Reality) realisiert werden: Freihand-Bedienung, integrierte Barcodescanner und die Reduzierung der Blickabwendungen auf externe Displays bieten bei der Arbeit mit Datenbrillen deutliche Vorteile gegenüber herkömmlichen Kommissioniersystemen.

Neben den genannten Einsatzszenarien in der Intralogistik finden sich weitere potenzielle Anwendungsfelder in allen Arbeitsbereichen, die über einen Bedarf an Anleitung, Informationsaufnahme, Kollaboration oder externer Kontrolle verfügen, bei denen aber Blickabwendungen oder das Halten und die Interaktion mit einem externen Display hinderlich sind.

Die Einblendung für die Tätigkeit relevanter, kontextsensitiver Informationen zu verschiedenen Arbeitsschritten kann beispielsweise zu einer Reduzierung des Schulungsaufwands (Training on the job) führen oder eine effizientere Arbeitsverrichtung durch schrittweise Arbeitsanleitungen gewährleisten. Außerdem lassen sich für einzelne Arbeitsschritte auch Zweitmeinungen (Remote Expert) einholen oder die Arbeitsqualität per Videostream und -dokumentation überprüfen [2].

1.3 Laborstudie von BGHW und IFA: Anzeigesysteme auf Gabelstaplern

Allein im Jahr 2016 meldete die Deutsche Gesetzliche Unfallversicherung (DGUV) 218 213 Arbeitsunfälle im innerbetrieblichen Verkehr, wobei die meisten Unfälle mit Flurförderzeugen (Gabelstapler, Handwagen) und 12 671 mit Gabelstaplern auftraten [3]. Die häufigste Unfallursache waren Fehlverhalten beim Fahren oder unzureichende betriebliche Sicherheitsvorkehrungen. Und obwohl es keine genauen Zahlen über ablenkungsbedingte Unfälle gibt, zeigen entsprechende Studien zum Einsatz von Informations- und Kommunikationstechnologien im Straßenverkehr, dass die Reduzierung von fahrfremden Tätigkeiten einen wesentlichen Beitrag zur Unfallverhütung leisten kann [4]. Insofern sollten diese

Arbeitsplätze so gestaltet sein, dass die kognitive Belastung für die Beschäftigten minimiert wird [5].

Trägt man eine Datenbrille, dann werden Informationen in der Ego-Perspektive präsentiert, was zu einer Reduzierung der blickbezogenen Ablenkung von der Fahrtätigkeit führen kann. Daher ist zu prüfen, ob dieser positive Aspekt mit einer Reduzierung der kognitiven Arbeitsbelastung einhergeht. Untersucht wird sowohl der Unterschied zwischen dem Einsatz von Displays gegenüber Fahrtätigkeiten ohne Displays als auch, ob und inwieweit sich die kognitive Belastung von Beschäftigten durch den Einsatz von Datenbrillen auf Gabelstaplern gegenüber herkömmlichen Monitorsystemen unterscheidet. Untersucht wurde der Einsatz mit zwei Datenbrillen (monokular und binokular) und einem Monitor (Tablet).

1.4 Forschungsfrage

In der vorliegenden Studie wird untersucht, ob durch das Benutzen von Displays auf Gabelstaplern während der Fahrt eine kognitive Belastung für die Beschäftigten besteht und wie hoch diese ist. Zudem wird untersucht, ob ein Unterschied in der kognitiven Belastung zwischen den untersuchten Displays – Tablet, monokulare sowie binokulare Datenbrille – festzustellen ist.

2 Methodik

2.1 Studienkollektiv – Beschreibung

Es wurden nur Angestellte der beiden Projektpartner Fa. CINOVAION und IFA als Probanden in das Studienkollektiv aufgenommen. Dazu war es notwendig, dass die Probanden einen Ausbildungsnachweis nach dem DGUV Grundsatz 308-001 [6] besaßen. Außerdem durften die Fahrer keine Brille tragen oder mussten auch ohne Brille Gabelstapler fahren können. Diese Einschränkung war nötig, da die kopfgetragenen Anzeigeegeräte nicht zusammen mit Brillen benutzt werden konnten. Über die Rahmenbedingungen des Versuchs wurden die Probanden informiert (Anhang 1).

Ansonsten erfolgte die Einladung der Probanden nach betrieblichen Kriterien, vor allem danach, ob sie im Studienzeitraum verfügbar waren. Danach wurden die Probanden unterwiesen und sie bestätigten, dass sie freiwillig an der Studie teilnahmen (Anhang 1).

Insgesamt wurden 32 männliche Probanden eingeladen (22 Fa. CINOVAION, zehn IFA). Daten von neun Probanden wurden nicht ausgewertet, weil es zu Fällen von Übelkeit (Simulatorkrankheit) kam oder nur unvollständige Datensätze vorlagen.

Das Alter der verbleibenden 23 Probanden lag zwischen 23 und 53 Jahren, der Mittelwert bei 40 Jahren. Ein Proband trug zwar eine Brille, nahm aber ohne Brille an dem Versuch teil. Die anderen 22 Probanden waren keine Brillenträger.

Schließlich ergab die Selbstauskunft der Probanden, dass die Benutzungsdauer von Gabelstaplern im Monat zwischen 0,3 und 40,0 h lag. Der Mittelwert lag bei 6,5 h und der Median bei 3,0 h.

3 Versuchsablauf

Um die Belastung durch den Einsatz von Datenbrillen objektiv zu beurteilen, wurden Reaktionszeiten und Fehlerraten nach ISO 17488 [7] erhoben. Darüber hinaus mussten die Probanden Fragebögen zur subjektiven Belastung ausfüllen.

3.1 Messung der Reaktionszeiten

Die drei Aufgaben nach ISO 17488 wurden wie folgt umgesetzt.

- Als Primäraufgabe diente das Fahren im Fahrsimulator, der ein virtuelles Lager ohne anderen Verkehr mit nummerierten Regaltrassen anzeigte. In diesem Lager mussten Regale mit bestimmten Nummern angefahren werden und es wurden Fahrfehler in jeder Fahrt gezählt. Ein Fahrfehler lag vor, wenn das Fahrzeug in der Simulation Gegenstände (Wände, Regale) berührte.

- Die Sekundäraufgabe bestand im Auslesen von Informationen aus drei verschiedenen Arten von Anzeigegeräten. Die Informationen bestanden entweder aus der Angabe einer zufällig ausgewählten Regalnummer, die angefahren werden musste, und aus eingblendeten, einstelligen Zahlen, die vorgelesen werden mussten und alle fünf Sekunden wechselten. Außerdem wurden Anfang und Ende einer Fahrt angezeigt. Das Erreichen des Zielregals musste quittiert werden. Die Vergleichsaufgabe bestand im Fahren ohne Anzeigegeräte, wobei die Versuchsleitung die anzufahrenden Regalnummern vorgelesen hat.
- Der Reaktionstest wurde nach Norm mit einem optischen Stimulus durchgeführt. Die Leuchtdiode befand sich im Sichtfeld der Probanden (Abbildung 1). Die Fahrten dauerten zwei Minuten, sodass die Fahrer auf etwa 30 Stimuli reagieren mussten. Erfolgte zwischen 100 und 2 500 ms eine Reaktion, so wurde dies als „Treffer“ gezählt und die jeweilige Reaktionszeit gespeichert.

Abbildung 1:

Versuchsaufbau: HMD1 (oben links), HMD2 mit externem Touchpad (oben rechts), Tablet (unten links), Position der Leuchtdiode (unten rechts)



Durch die nach Norm durchgeführte Übungsphase wurde sichergestellt, dass alle Probanden die Primäraufgabe gut bewältigen konnten. Dadurch wurden auch Unterschiede ausgeglichen, die zwischen realen Fahrzeugen und dem Simulator bestanden, z. B. das unterschiedliche Verhalten der Lenkung. Nach der Übungsphase absolvierten die Probanden vier Fahrten:

- ohne Anzeigegerät („Baseline“),
- mit monokularem Anzeigegerät (head-mounted display 1 „HMD1“),
- mit binokularem Anzeigegerät (head-mounted display 2 „HMD2“),
- mit einem externen Monitor („Tablet“).

Die Reihenfolge der Fahrten wurde jeweils zufällig festgelegt. Da das Versuchsziel darin bestand, die Belastung durch Anzeigeräte zu bestimmen und eventuell Unterschiede zwischen ihnen festzustellen, wurde ein optischer und kein taktiler Stimulus gewählt, um einen höheren Effekt in den Leistungsparametern zu erzeugen. Außerdem wurde durch das Vorlesen von Zahlen erreicht, dass während des Versuchs eine kontinuierliche Belastung in der Sekundäraufgabe vorlag. So wurde erst ein sinnvoller Vergleich mit der Fahrt ohne Anzeigerät möglich.

3.2 Gabelstaplersimulator

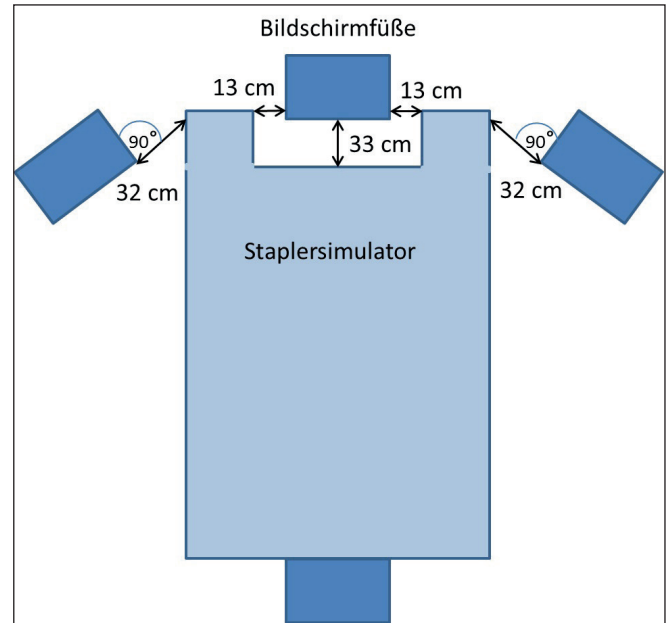
Für die Studie wurde der Gabelstaplersimulator des IFA verwendet. Die reale Arbeitsumgebung erscheint als 3D-Simulation, das Fahrzeug bietet einen haptischen Fahrgastraum (Abbildung 2). Die Simulationssoftware (Unity Engine) wurde auf drei 55-Zoll-Monitoren in Full-HD-Qualität für die Frontsicht und einem Bildschirm für die Rücksicht dargestellt. Je nach Einstellung des lageverstellbaren Fahrersitzes variierte der Betrachtungsabstand zwischen 85 und 100 cm (Abbildung 3). Sowohl der technische Aufbau als auch deren Anordnung basierten auf einem maximal erreichbaren Realitätsgrad der betrieblichen Praxis.

Um eine realistischere Simulation zu erreichen, wurde die Simulation während der Studienlaufzeit um einen eingblendeten Mast ergänzt. Diese ergänzte Simulation wurde von neun Probanden aus beiden beteiligten Betrieben durchlaufen. Wegen der in Abschnitt 2.1 „Studienkollektiv – Beschreibung“ genannten Gründe konnten nur die Daten von fünf Probanden ausgewertet werden. Da sich diese nicht hinreichend von den anderen Ergebnissen unterschieden, wurden die Ergebnisse beider Simulationen zusammengefasst (siehe Abschnitt 4.5 „Ergebnisse – Masteinblendung“).

Abbildung 2:
Gabelstaplersimulator



Abbildung 3:
Anordnung des Gabelstaplersimulators mit Monitoren

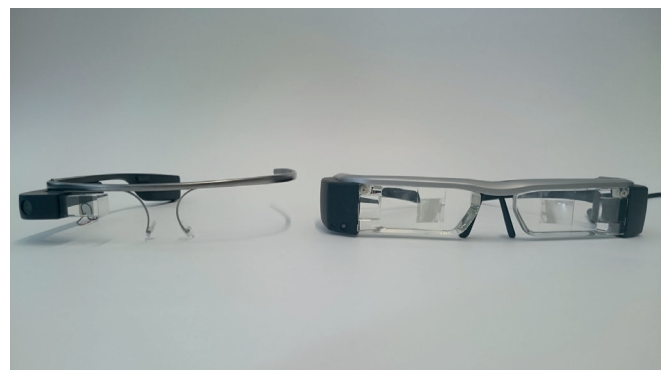


3.3 Displays

Bei der Studie wurden drei verschiedene Displays untersucht: Ein 10,1-Zoll-Tablet (Tablet), das in Bedienung, Maßen und Anbringung mit einem in der Praxis auf Gabelstaplern verwendeten Monitorsystem vergleichbar ist. Für die Versuche mit dem Tablet wurde das Gerät mit einer Halterung an der Mittelkonsole befestigt (Abbildung 1 unten links).

Für die Versuche mit den Datenbrillen wurden zwei Geräte unterschiedlicher Bauart verwendet: eine Datenbrille (HMD1) mit monokularem Display und Touchpad an der rechten Seite des Brillengestells und eine Datenbrille (HMD2) mit binokularem Anzeigesystem und einem externen Touchpad, das über ein Kabel mit dem Rahmen verbunden und in der Mittelkonsole montiert wurde (Abbildung 1 oben rechts). Die Datenbrillen sind in Abbildung 4 dargestellt.

Abbildung 4:
HMD1 (links) und HMD2 (rechts)



3.4 Fragebögen

Die vollständigen Fragebögen (siehe Anhang 2) bestanden aus einem allgemeinen Teil, der Fragen zu folgenden Punkten enthielt:

- Alter in Jahren
- Geschlecht, Brillenträger (ja/nein)
- Fahrpraxis in Stunden pro Monat
- abgelegte Prüfung nach [6].

Zusätzlich wurde nach jeder Fahrt die subjektive Belastung mithilfe eines standardisierten Fragebogens abgefragt, den Beschäftigte der amerikanischen Weltraumbehörde entwickelten und der es ermöglicht, Arbeitsbelastungen zu bewerten (englisch: task load index (TLX)) [8]. Daher wird dieser Fragebogen in der Literatur mit NASA-TLX abgekürzt.

Er besteht aus sechs Einzelfragen zu unterschiedlichen Aspekten der Belastung, die durch das Setzen eines Kreuzes auf einer Skala beantwortet wurden. Da die Skala 20 Teilstriche zwischen „gering/gut“ und „hoch/schlecht“ enthielt, wurden die Kreuze als Zahl zwischen 0 und 20 ausgewertet. Dabei wurde auf eine Gewichtung der Einzelfragen verzichtet (siehe Abschnitt 4.4 „Ergebnisse – NASA-TLX“).

Schließlich wurden allgemeine Fragen danach gestellt, welches System am besten gefiel, welches am wenigsten abgelenkt hatte und ob ein Anzeigesystem die Sicht behinderte.

4 Ergebnisse

4.1 Ergebnisse – Allgemein

Aus dem Versuchsaufbau stehen drei Leistungsparameter aus den Messungen zur Verfügung: die Fahrfehler, die Trefferrate (Anzahl Treffer dividiert durch die Anzahl der Stimuli) und die Reaktionszeiten der Treffer (mittlere Reaktionszeit).

Hier ist nach Norm vor allem die mittlere Reaktionszeit als Maß für die sensorische oder kognitive Belastung wichtig. Die Trefferrate kann aber ebenfalls herangezogen werden. Die Zahl der Fahrfehler ist in der Norm nicht als Leistungsparameter vorgesehen und zunächst ist zu klären, ob sie für diese Versuchsreihe Aussagen über die Belastung durch die Anzeigergeräte beziehungsweise über das Zusammenwirken von Probanden und Fahrsimulator möglich macht. Außerdem sollen die Auswirkungen der Masteinblendung analysiert werden.

Bei den Fragebögen kommt dem NASA-TLX als standardisiertem Fragebogen die zentrale Rolle zu. Die weiteren allgemeinen Fragen können nur unterstützende Informationen liefern.

Da für jeden Probanden Messungen und Fragebögen zu den drei Anzeigergeräten vorliegen, sind diese Ergebnisse nicht unabhängig voneinander, sodass einfache Varianzanalysen nicht möglich sind. Eine Möglichkeit, dieses Problem zu umgehen, ist es, Differenzen der Maßzahlen für die einzelnen Probanden zu bilden, und einem t-Test zu unterziehen (z. B. Abschnitt 4.3 „Ergebnisse – Reaktionstest“) [9].

4.2 Ergebnisse – Fahrfehler

Die Fahrfehler werden automatisch von der Simulationssoftware erfasst und können somit ausgewertet werden, obwohl sie nicht Teil der Versuchsanordnung nach ISO 17488 sind.

In Tabelle 1 sind die Fahrfehler für die 23 Probanden abhängig von der Nutzung der Anzeigergeräte aufgetragen. 71 der 92 Fahrten sind frei von Fahrfehlern. Vereinzelt kommen ein oder zwei Fehler vor und nur bei drei Fahrten sind vier Fahrfehler zu verzeichnen. Summiert man über Anzeigergeräte auf, so sind für die Vergleichsfahrt vier und für HMD2 und Tablet je zwölf Fahrfehler zu verzeichnen. Allerdings kommen bei der Nutzung von HMD1 nur fünf Fahrfehler vor.

Das wirft die Frage auf, ob nachgewiesen werden kann, dass die Fahrfehler abhängig sind von den Probanden und der Gerätenutzung, oder ob sie von einer Situation nicht unterscheidbar sind, bei der die Fahrfehler rein

zufällig zustande kommen. Die Null-Hypothese ist demnach, dass die Anforderungen durch den Versuchsaufbau (kein weiterer Verkehr, ausgebildete Fahrer) und durch die Anzeigergeräte so gering sind, dass die Fahrfehler selten und voneinander unabhängig sind. Die Alternativ-Hypothese dazu ist, dass es gute und schlechte Fahrer gibt, die generell mehr oder weniger Fahrfehler machen. Außerdem sollten die Fahrfehler mit steigender kognitiver und sensorischer Belastung zunehmen.

Tabelle 1:
Fahrfehler; Nr. ist eine fortlaufende Nummer des Probanden, eingetragen sind die Fahrfehler für die Vergleichsfahrt (Baseline) und die drei Anzeigergeräte (siehe Seite 15, Abschnitt 3.1)

Nr.	Baseline	HMD1	HMD2	Tablet
1	1	0	0	0
2	0	1	0	0
3	0	0	0	1
4	0	4	0	0
5	0	0	0	0
6	2	0	1	2
7	0	0	0	0
8	0	0	0	0
9	0	0	1	1
10	1	0	0	0
11	0	0	0	0
12	0	0	0	1
13	0	0	0	0
14	0	0	0	0
15	0	1	0	0
16	0	0	0	1
17	0	1	0	4
18	0	0	0	0
19	0	0	1	1
20	0	0	0	1
21	0	4	2	0
22	0	1	0	0
23	0	0	0	0

Folgt man der Null-Hypothese, so legt das Histogramm der Fahrfehler in Abbildung 5 nahe, dass es sich um eine Poisson-Verteilung handeln könnte. Der Maximum-Likelihood-Schätzer für die Eintrittswahrscheinlichkeit der Fahrfehler ist

$$\lambda = \frac{33 \text{ Fehler}}{92 \text{ Fahrten}}$$

Um festzustellen, ob die Verteilung dieser Stichprobe einer Poisson-Verteilung entspricht, kann aber kein verlässlicher χ^2 -Test der Anpassungsgüte durchgeführt werden. Es können nur drei Kategorien mit mehr als fünf Beobachtungen gebildet werden (1. Kategorie: keine Fehler, 2. Kategorie: ein Fehler, 3. Kategorie: zwei bis

vier Fehler), was zu $v = 3 - 1 - 1 = 1$ Freiheitsgraden der χ^2 -Verteilung führt, da auch der Parameter λ der Poisson-Verteilung geschätzt werden muss. Es ergibt sich ein $\chi^2 = 3,89$ und damit $p = 0,048$, was nicht klar dafür spricht, dass das nötige Signifikanzniveau von $\alpha = 0,05$ erreicht wird.

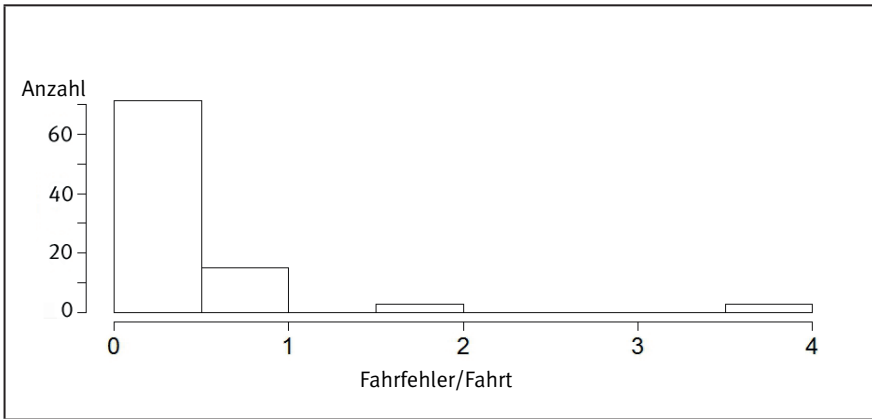


Abbildung 5:
Histogramm der Fahrfehler pro Fahrt

Die Null-Hypothese, dass Fahrfehler unabhängig sind, lässt sich auch mit einer Bootstrap-Simulation untersuchen. Dazu werden aus den 92 Ergebnissen in Tabelle 1 unabhängig mit Zurücklegen Zufallsstichproben von je 23 Ergebnissen gezogen, und deren Fahrfehler aufsummiert. Das Histogramm für 10^5 solcher Stichproben ist in Abbildung 6 zu sehen.

Die Stichproben in Abbildung 6 stammen aus einem rein zufälligen Prozess, wobei die Ergebnisse unabhängig voneinander gezogen werden. Die extremsten Werte, die in Tabelle 1 aufgeführt sind, betragen vier Fahrfehler pro

23 Fahrten (Baseline) und 12 Fahrfehler pro 23 Fahrten (HMD1, Baseline). Kommen diese oder noch extremere Werte mit einem Anteil von mehr als $\alpha = 0,05$ vor, so kann man die Null-Hypothese unabhängiger, zufälliger Fahrfehler nicht ablehnen.

Tatsächlich beträgt in Abbildung 6 der Anteil für vier und weniger Fahrfehler pro 23 Fahrten $p = 0,16$, und der Anteil für zwölf und mehr Fahrfehler $p = 0,20$. Das bedeutet, dass wir aus statistischer Sicht nicht genügend Argumente haben anzunehmen, dass die beobachteten Fahrfehler anders als rein zufällig zustande gekommen sind.

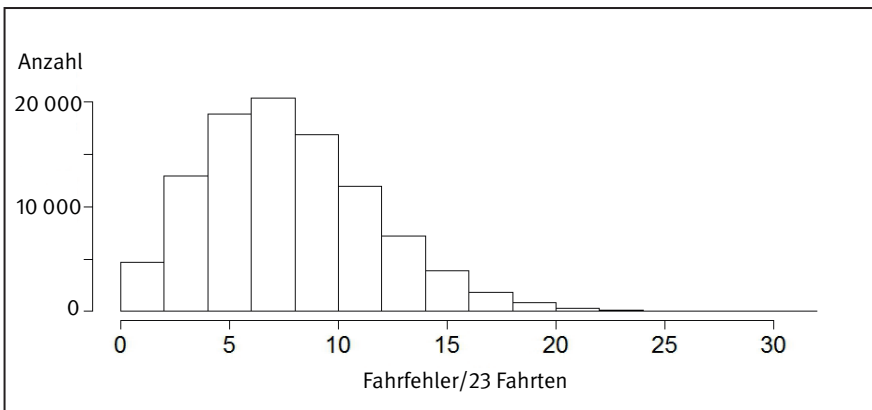


Abbildung 6:
Histogramm der Fahrfehler von 10^5 Bootstrap-Stichproben mit 23 Ereignissen aus Tabelle 1

4.3 Ergebnisse – Reaktionstest

4.3.1 Ergebnisse – Reaktionstest – Allgemein

Aufgrund von physiologischen Überlegungen geht die Versuchsnorm davon aus, dass die Verteilung der Reaktionszeiten einer logarithmischen Normalverteilung entspricht, und nur solche Datensätze sollen in die Auswertung einfließen.

Dadurch, dass eine gültige Reaktionszeit nur in dem physiologisch sinnvollen Abschnitt zwischen 100 und 2 500 ms nach dem Stimulus aufgenommen wird, sind die Voraussetzungen gut, dass die Reaktionszeiten auch erwartungsgemäß verteilt sind. Daher wird das Vorhandensein einer logarithmischen Normalverteilung nicht vollständig überprüft, aber es wird ein Plausibilitätstest durchgeführt, um festzustellen, dass keine großen Fehler bei der Datenaufnahme vorliegen.

Dazu werden in den Datensätzen jeder Fahrt von allen Reaktionszeiten 100 ms abgezogen, damit der Ursprung der Verteilung beim Wert Null liegt. Dann wird der Mittelwert des logarithmierten Datensatzes mit dem logarithmierten Median des ursprünglichen Datensatzes verglichen, die bei logarithmischen Normalverteilungen gleich sein müssen [11]. Bei den vorliegenden Datensätzen liegt dieser Unterschied unter 1 ms und wir gehen davon aus, dass die Datensatzstruktur korrekt ist.

4.3.2 Ergebnisse – Reaktionstest – Belastungszunahme bei Gerätenutzung

In Abbildung 7 sind die Mittelwerte der Reaktionszeiten für jeden Probanden als Punktwolke (blaue Kreise) in vier Spalten eingetragen, die die unterschiedlichen Geräte-nutzungen darstellen (siehe Abschnitt 3.1 „Messung der Reaktionszeiten“).

Zusätzlich sind Boxplots eingetragen, die die Verteilung der Punktwolke beschreiben:

- Die oberen und unteren Ränder des Kastens beschreiben das 75. und 25. Perzentil der Stichprobe.
- Der Strich innerhalb des Kastens zeigt den Median an.
- Die Ausläufer an dem Kasten reichen zu extremsten Messwert, der nicht weiter als das 1,5-Fache des Interquartilabstandes (Betrag des Abstands zwischen dem 25. und 75. Perzentil) von den Kastenrändern entfernt ist.
- Werte, die weiter von den Kastenrändern entfernt liegen, werden als ungefüllte Kreise dargestellt.

In der Abbildung ist zu sehen, dass die vier Verteilungen nicht symmetrisch sind und überlappen, auch wenn die Fahrten ohne Gerätenutzung (Baseline) in der Tendenz niedrigere Werte liefern.

Um von dieser rein deskriptiven Betrachtung zu einer quantitativen Aussage zu kommen, muss man sich vor Augen führen, dass man es mit einer verbundenen Stichprobe zu tun hat (siehe Abschnitt 4.1 „Ergebnisse – Allgemein“). Um dies zu veranschaulichen, sind in Abbildung 7 Pfeile eingetragen, die die mittleren Reaktionszeiten eines Probanden bei den vier Gerätenutzungen verbinden.

Es ist zu sehen, dass hohe mittlere Reaktionszeiten bei der Fahrt ohne Geräte auch dazu führen, dass die Reaktionszeiten bei den Fahrten mit Geräten hoch sind. Außerdem ist nicht immer die Fahrt ohne Gerät mit der geringsten Reaktionszeit verbunden.

Um zu zeigen, dass die Pfeile in Abbildung 7 keinen Einzelfall beschreiben, sind in Abbildung 8 die Werte aus Abbildung 7 für jeden Probanden und jeweils zwei Nutzungsarten gegeneinander aufgetragen. Dabei sind lineare Zusammenhänge zwischen einzelnen Nutzungsarten offensichtlich, die es nicht möglich machen, einfache varianzanalytische Verfahren anzuwenden.

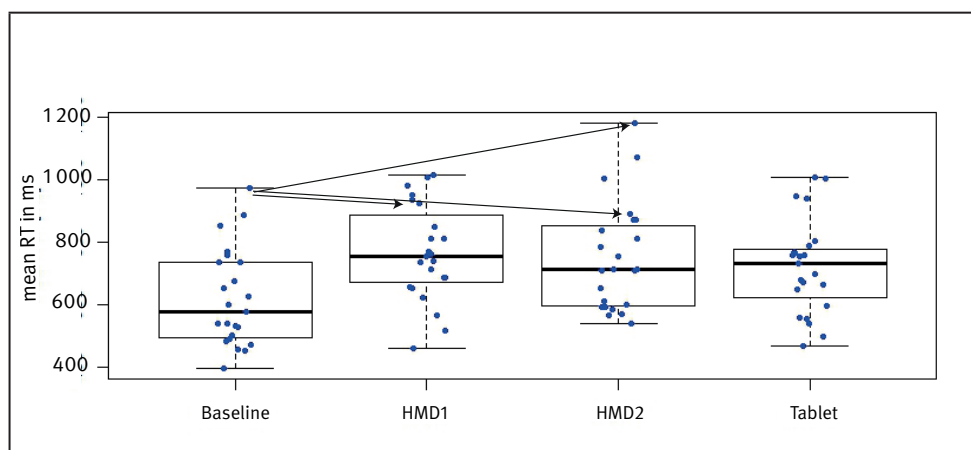


Abbildung 7: Verteilung der mittleren Reaktionszeiten pro Proband als Punktwolke (blaue Kreise) und Boxplot für vier verschiedene Arten von Fahrten. Pfeile verbinden die mittleren Reaktionszeiten eines Probanden.

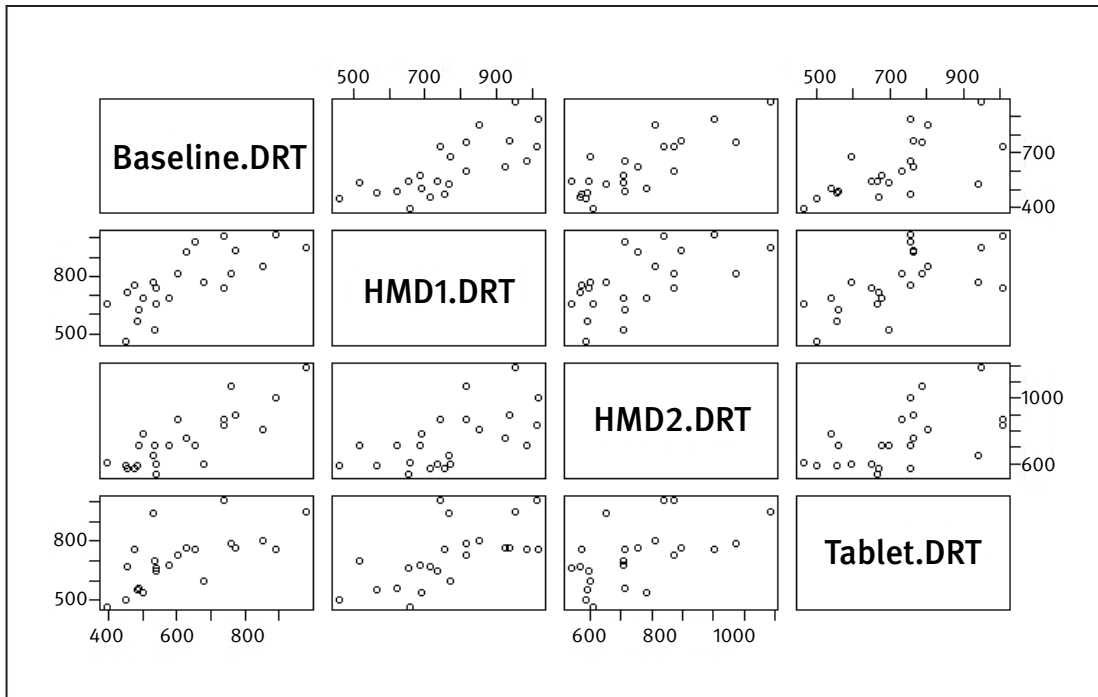


Abbildung 8:
Abhängigkeit der
Reaktionszeiten;
Werte aus Abbil-
dung 7 pro Proband
aufgetragen für je
zwei Nutzungsarten

Deswegen werden, wie in Abschnitt 4.1 angedeutet, die Differenzen der mittleren Reaktionszeiten pro Proband ausgewertet [9]. In Abbildung 9 sind die Differenzen pro Proband mit Bezug zu den Fahrten ohne Gerätenutzung (Baseline) wieder als Punktwolke und als Boxplot aufgetragen. Zunächst ist festzuhalten, dass die Verteilungen stark überlappen und eine gewisse Schiefe aufweisen. Außerdem liegt der Hauptteil der Verteilungen bei negativen Werten (75. Perzentils negativ), was anzeigt, dass die mittlere Reaktionszeit pro Proband bei der Gerätenutzung zugenommen hat. Positive Differenzen und damit eine Abnahme der Belastung bei Gerätenutzung sind die Ausnahme. Im Mittel nehmen die Reaktionszeiten um 147 ms (HMD1), 130 ms (HMD2) und 103 ms (Tablet) ab.

Unter der Null-Hypothese, dass die Abnahme rein zufällig ist, müssten diese Mittelwerte sich nicht signifikant vom Wert Null unterscheiden. Dies kann mit einem t-Test überprüft werden. Aber dazu muss zunächst festgestellt

werden, ob die Stichproben der Differenzen einer Normalverteilung entsprechen. Wegen der Schiefe der Daten ist dies aus Abbildung 9 nicht offensichtlich und daher wird jeder Datensatz einem Shapiro-Wilk-Test [11] unterzogen. Bei diesem Test ist es wichtig, das Signifikanzniveau nicht zu gering zu setzen, da die Null-Hypothese ist, dass die Daten normalverteilt sind. Die Signifikanzwerte liegen für die hier gezeigten Daten mit $p = 0,24$ (Baseline – HMD1), $p = 0,48$ (Baseline – HMD2) und $p = 0,82$ (Baseline – Tablet) von einem Standardwert $\alpha = 0,05$ weit genug entfernt, um von Normalverteilungen ausgehen zu können.

Da die Voraussetzung für den t-Test somit gegeben sind, muss nun noch das Signifikanzniveau festgelegt werden. Anhand dieses Datensatzes können sechs verschiedene Differenzen untersucht werden: drei mit Bezug zu Fahrten ohne Gerätenutzung (Abbildung 9) und drei im Vergleich der Gerätenutzung untereinander (Abbildung 10).

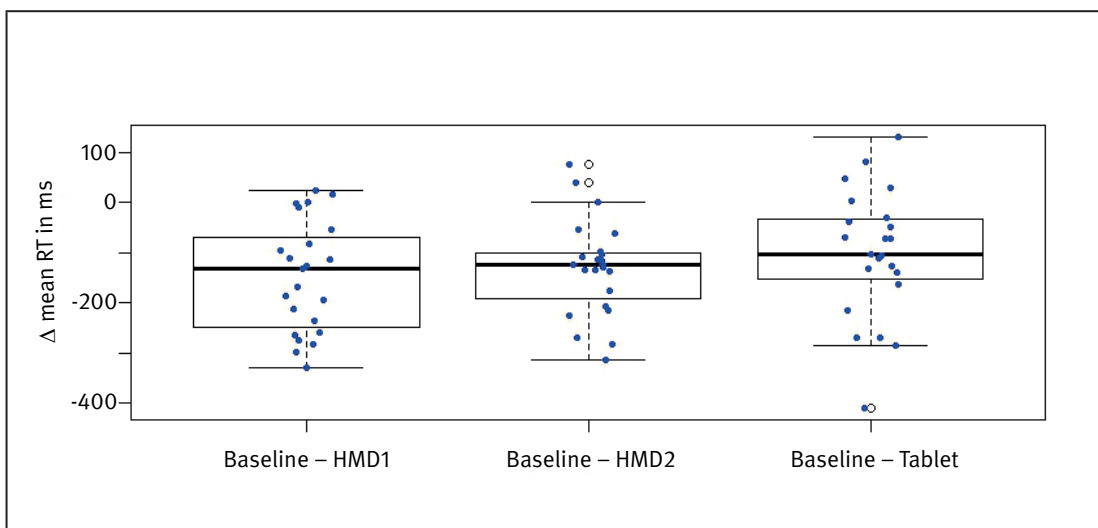


Abbildung 9:
Verteilung der Diffe-
renzen der mittleren
Reaktionszeiten pro
Proband, aufgetra-
gen als Punktwolke
(blaue Kreise) und
Boxplot (Bezug:
Baseline); Diffe-
renzen zwischen
Fahrten ohne und
mit Gerätenutzung

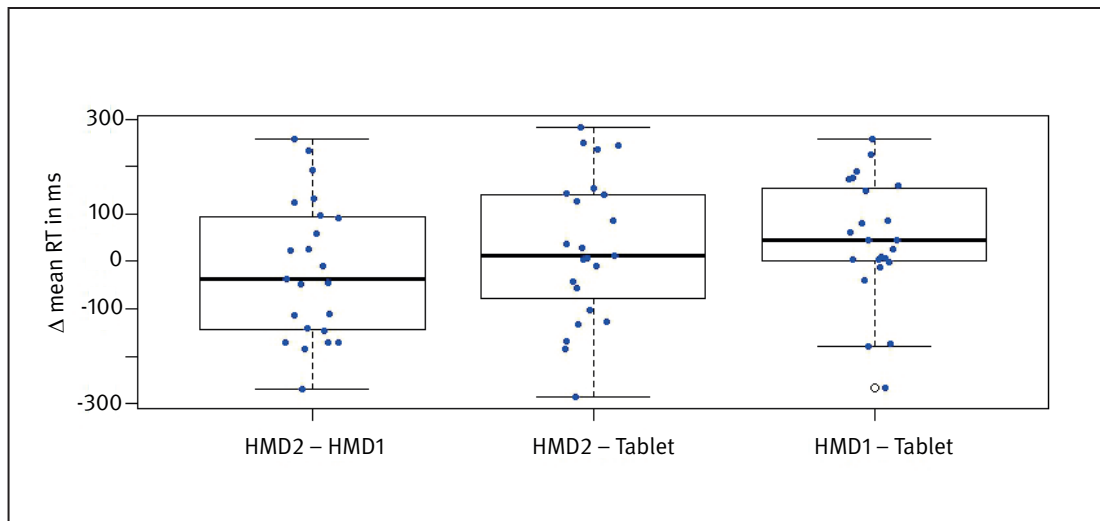


Abbildung 10: Verteilung der Differenzen der mittleren Reaktionszeiten pro Proband, aufgetragen als Punktwolke (blaue Kreise) und Boxplot (zwischen Geräten). Die Differenzen beziehen sich auf Fahrten mit unterschiedlichen Geräten. Minuend und Subtrahend entsprechen der Beschriftung.

Mehrfache Tests auf demselben Datensatz führen dazu, dass Null-Hypothesen öfter abgelehnt werden als bei einem Test. Man spricht daher auch von einer Kumulation von Fehlern 1. Art, dem die Bonferroni-Korrektur entgegenwirkt. Diese konservative Festlegung beruht darauf, dass nicht von einem gewünschten Signifikanzniveau (z. B. $\alpha=0,05$) ausgegangen wird, sondern dass dieses mit der Anzahl der Tests korrigiert wird. In unserem Fall ergibt sich das Signifikanzniveau der Tests zu

$$\alpha^* = \frac{\alpha}{6} \approx 0,0083$$

Für die Differenzen der mittleren Reaktionszeiten zwischen Fahrten ohne Gerätenutzung und mit Gerätenutzung unterscheiden sich in einem t-Test die Mittelwerte signifikant vom Wert Null: Im Ergebnis liegen die p-Werte unter 0,00085. Damit kann die Null-Hypothese für das korrigierte Signifikanzniveau α^* in allen drei Fällen abgelehnt werden.

Die Analyse weist damit darauf hin, dass mit dem Nutzen der Geräte eine Zunahme der sensorischen und kognitiven Belastung verbunden ist.

4.3.3 Ergebnisse – Reaktionstest – Belastungsunterschiede zwischen unterschiedlichen Geräten

Auf dieselbe Art und Weise, wie die Belastungszunahme bei der Gerätenutzung nachgewiesen wurde, kann man nun auch versuchen, Unterschiede zwischen den sensorischen und kognitiven Belastungen der verschiedenen Geräte zu bestimmen. Allerdings zeigt ein Blick auf Abbildung 10, dass es keine klare Abnahme oder Zunahme der Reaktionszeiten gibt. Der Wert Null liegt bei allen Verteilungen zwischen dem 25. und 75. Perzentil. Die Mittelwerte liegen bei -16 ms (HMD2 – HMD1), 28 ms (HMD2 – Tablet) und 44 ms (HMD1 – Tablet) und damit viel näher an dem Wert Null als bei den Unterschieden im Vergleich zu den Fahrten ohne Gerätenutzung.

Zwar erfüllen diese drei Stichproben den Shapiro-Wilk-Test mit einem Signifikanzwert deutlich über 0,05 ($p = 0,59$ HMD2 – HMD1, $p = 0,40$ HMD2 – Tablet, $p = 0,22$ HMD1 – Tablet), aber die folgenden t-Tests zeigen keinen signifikanten Unterschied zum Wert Null, weder auf dem Niveau $\alpha = 0,05$ noch auf dem korrigierten Niveau $\alpha^* = 0,0083$.

Neben den Reaktionszeiten empfiehlt die Prüfnorm auch die Untersuchung der Trefferraten (hit rates). Damit ist für jede Fahrt das Verhältnis aus der Anzahl der gültigen Reaktionen (Abschnitt 4.3.1) zur Gesamtzahl der Stimuli gemeint. In Abbildung 11 sind die Verteilungen der Trefferraten als Punktwolken und Boxplots für die unterschiedlichen Gerätenutzungen aufgetragen. Zu sehen ist dort, dass die Trefferraten tendenziell abnehmen, wenn Geräte während der Fahrt genutzt werden, was eine ähnliche Situation ist wie in Abbildung 7. Da die Analyse der Reaktionszeiten schon gezeigt hat, dass die kognitive und sensorische Belastung durch das Nutzen der Anzeigeräte zunimmt, soll es jetzt nur um eventuelle Unterschiede zwischen den Geräten gehen, die bei den Reaktionszeiten nicht nachweisbar gewesen sind.

Allerdings sind in Abbildung 11 keine großen Unterschiede in den Verteilungen bei den drei Gerätenutzungen zu sehen. Nach ISO 17488 sollten die Trefferraten mit einer logistischen Regression untersucht werden, aber das scheidet hier auch aus, weil wir es mit einer verbundenen Stichprobe zu tun haben und die Effekte der Gerätenutzung nicht unabhängig voneinander wären.

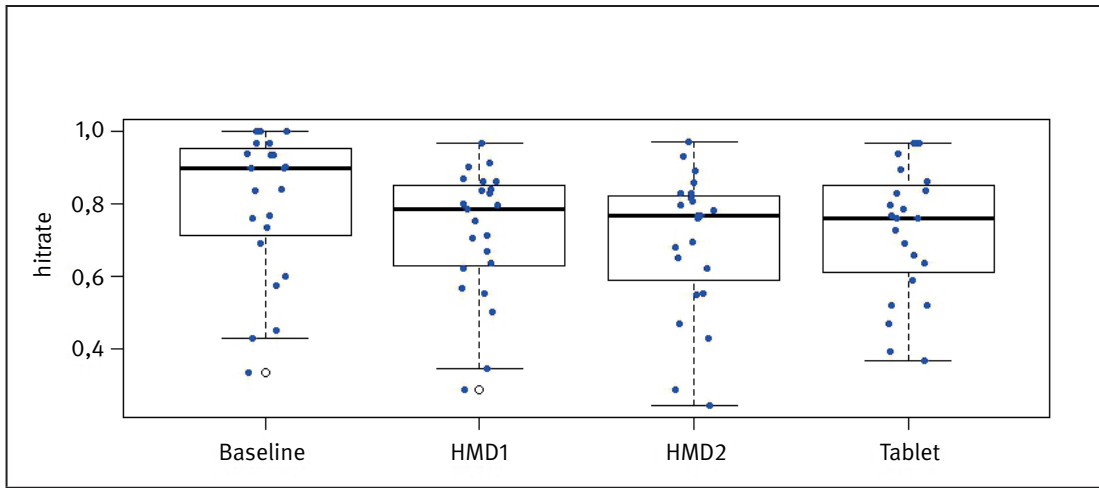


Abbildung 11: Verteilung der Trefferraten pro Proband als Punktwolke (blaue Kreise) und Boxplot für vier verschiedene Arten von Fahrten

Stattdessen wird zum Vergleich der Trefferraten bei unterschiedlicher Gerätenutzung eine numerische Simulation herangezogen, da die Stichproben für die einzelnen Probanden oft zu klein sind, um analytische Tests anwenden zu können [9]. Somit wird von der Null-Hypothese ausgegangen, dass die Treffer rein zufällig verteilt sind und nichts mit der Gerätenutzung zu tun haben. Kenngrößen sind die Differenzen der Trefferraten für unterschiedliche Geräte, also $p_{HMD2} - p_{HMD1}$, $p_{HMD2} - p_{Tablet}$ und $p_{HMD1} - p_{Tablet}$. Unter der Null-Hypothese sind alle diese Differenzen gleich Null.

Es wird nun für jeden Probanden überprüft, ob die gefundenen Unterschiede zwischen den Trefferraten auf zufällige Abweichungen zurückzuführen sind oder nicht. Tabelle 2 enthält ein Beispiel für den Unterschied zwischen HMD2 und Tablet eines Probanden:

$$p_{HMD2} - p_{Tablet} = \frac{19}{28} - \frac{11}{30} \approx 0,312$$

Tabelle 2: Vergleich Trefferraten für Proband 16

Proband 16	HMD2	Tablet	Summe
Treffer	19	11	30
Fehler	9	19	28
Summe	28	30	58

Nun werden 100 000 Stichproben gebildet, in denen 30 Treffer und 28 Fehler zufällig so auf die vier Felder verteilt werden, dass wieder 28 Versuchsergebnisse HMD2 zugeordnet werden und 30 der Tabletnutzung. Für jede dieser Stichproben wird die Differenz $p_{HMD2} - p_{Tablet}$ gebildet und diejenigen Differenzen gezählt, deren Betrag über dem Betrag des experimentellen Ergebnisses liegt (zweiseitiger Test). Dies war bei 2,1 % der Stichproben der Fall, was unterhalb des Signifikanzniveaus von $\alpha = 0,05$ liegt. Für den Proband 16 ist damit die Tabletnutzung signifikant schlechter als die Nutzung von HMD2.

Wiederholt man dies für alle Probanden und alle verschiedenen Gerätenutzungen, stellt man aber nur bei drei Probanden einen signifikanten Unterschied im Vergleich unterschiedlicher Geräte fest: Proband 10 ($p_{HMD2} - p_{HMD1}$), Proband 16 ($p_{HMD2} - p_{Tablet}$) und Proband 30 ($p_{HMD2} - p_{Tablet}$, $p_{HMD1} - p_{Tablet}$). Bei den anderen 20 Probanden hat es beim Nutzen der drei verschiedenen Geräte keine Änderungen in den Trefferraten gegeben, die von rein zufälligen Änderungen auf dem Signifikanzniveau $\alpha = 0,05$ unterschieden werden können.

4.4 Ergebnisse – NASA-TLX

4.4.1 Ergebnisse – NASA-TLX – Belastungszunahme bei Gerätenutzung

Im Folgenden stellt der Mittelwert der Angaben zu den sechs Fragen pro Proband und Fahrt die Kennzahl der Belastung dar (TLX-Mittelwert). Der TLX-Mittelwert kann nur Werte zwischen Null und 20 annehmen (Abschnitt 3.4 „Fragebögen“).

Mit dem Nutzen des Mittelwertes wird darauf verzichtet, die Belastungsaspekte unterschiedlich zu gewichten. Diese Mittelwertbildung wird in der Literatur als „raw NASA-TLX“ bezeichnet. Somit bezieht sich die Analyse dieses Abschnitts auf eine allgemeine Belastung und nicht auf die kognitive und sensorische Belastung, die der Reaktionstest nach ISO 17488 untersucht. Eine genauere Begründung für dieses Vorgehen wird im Abschnitt Ergebnisse-NASA-TLX-Einzelfragen gegeben.

In Abbildung 12 sind die Verteilungen der TLX-Mittelwerte als Punktwolke und Boxplot aufgetragen. Wie bei den Reaktionszeiten (Abbildung 7) ist auch hier tendenziell eine geringere Belastung bei den Fahrten ohne Geräte zu sehen, während die Verteilungen für die Fahrten mit Gerätenutzung in der Tendenz bei höheren Werten liegen und sich stark überlappen.

Da auch diese Daten aus verbundenen Stichproben stammen, muss in einem nächsten Schritt untersucht werden, wie sich die Differenzen der Belastungen darstellen. Dabei wird, wie bei den Reaktionszeiten, pro Proband die Belastungszahl der Gerätenutzung von der Belastungszahl ohne Gerätenutzung abgezogen. Damit sind wieder negative Differenzen mit einer Belastungszunahme verbunden.

In Abbildung 13 ist zu sehen, dass die Fahrten mit Gerät mit einer höheren Belastung verbunden sind. Im Mittel nehmen die TLX-Mittelwerte bei den Probanden um 3,5 (HMD1), 4,1 (HMD2) und 2,4 (Tablet) auf der TLX-Skala zu. Belastungsabnahmen, die positiven Werten entsprechen, sind die Ausnahme.

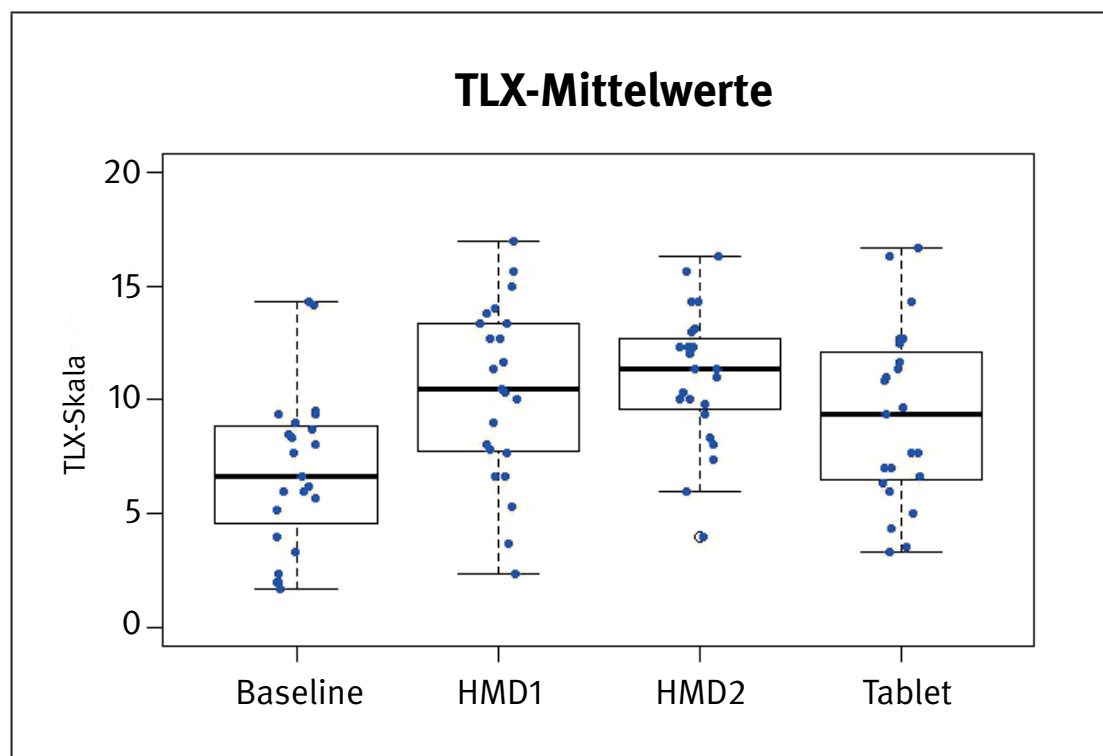


Abbildung 12: Verteilung der TLX-Mittelwerte pro Proband als Punktwolke (blaue Kreise) und Boxplot für vier verschiedene Arten von Fahrten

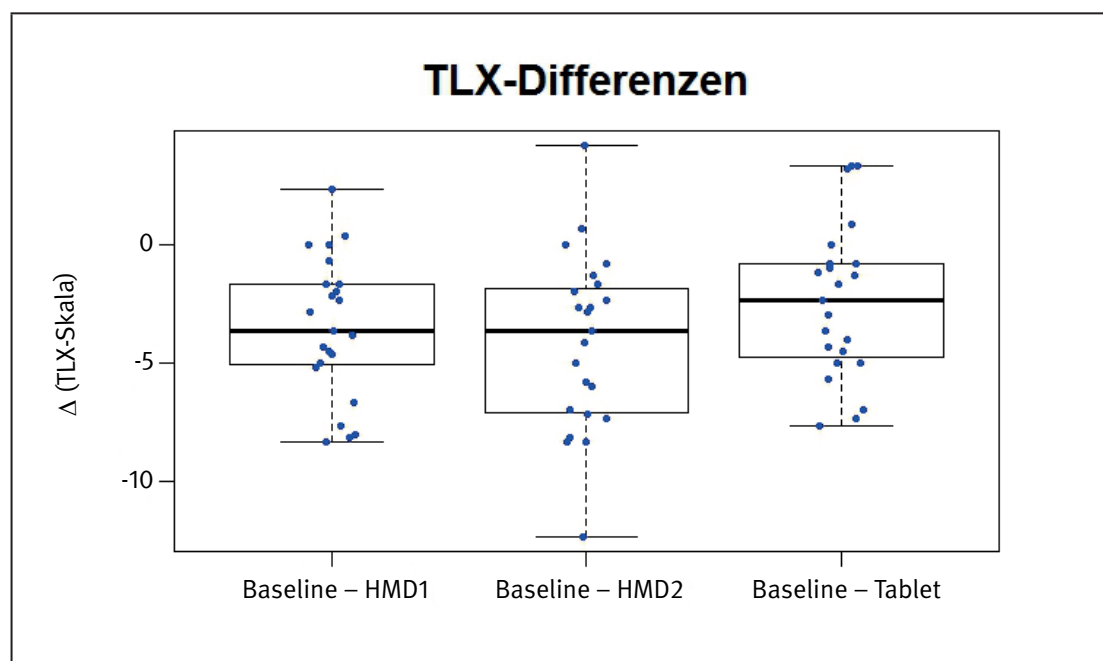


Abbildung 13: Verteilung der Differenzen der TLX-Mittelwerte pro Proband als Punktwolke (blaue Kreise) und Boxplot (Bezug: Baseline); Differenzen zwischen Fahrten ohne und mit Gerätenutzung

Als nächstes ist zu fragen, ob diese Zunahme der TLX-Mittelwerte signifikant ist. Dazu wird dasselbe Verfahren angewendet, wie bei den Reaktionszeiten. Zunächst werden Differenzen der Belastungswerte gebildet und die drei Verteilungen mit Bezug zur Fahrt ohne Anzeigergeräte

aus einem Shapiro-Wilk-Test [11] unterworfen, um zu überprüfen, ob diese einer Normalverteilung entsprechen (Abbildung 13). Die Signifikanzwerte liegen mit $p = 0,56$ (Baseline - HMD1), $p = 0,92$ (Baseline - HMD2) und $p = 0,39$ (Baseline - Tablet) weit genug von dem

Standardwert $\alpha = 0,05$ entfernt, um von Normalverteilungen ausgehen zu können.

Somit kann mit einem t-Test die Null-Hypothese getestet werden, dass der Mittelwert der Differenzen Null ist. Da wie bei den Reaktionszeiten sechs verschiedene Tests möglich sind, wird nach *Bonferroni* ein korrigiertes Signifikanzniveau von

$$\alpha^* = \frac{\alpha}{6} \approx 0,0083$$

verlangt. Die Tests ergeben, dass die TLX-Mittelwerte für die Verteilungen in Abbildung 13 mit einem p-Wert unter 0,0017 vom Wert Null verschieden sind und die Null-Hypothese für alle drei Fälle auf dem korrigierten Signifikanzniveau α^* abgelehnt werden kann.

Vergleicht man also Fahrten ohne Geräte und Fahrten mit Geräten, so ergibt die Analyse der Ergebnisse des NASA-TLX, dass eine allgemeine Belastung beim Nutzen der Geräte zunimmt.

4.4.2 Ergebnisse – NASA-TLX – Belastungsunterschiede zwischen unterschiedlichen Geräten

Analog zur Analyse der Reaktionszeiten kann nun auch die Analyse der TLX-Mittelwerte genutzt werden, um mögliche Unterschiede bei der Nutzung der verschiedenen Geräte festzustellen.

Abbildung 14 zeigt, dass der Wert Null bei allen Verteilungen von Belastungszahldifferenzen bei verschiedener Gerätenutzung zwischen dem 25. und 75. Perzentil liegt. Die Mittelwerte liegen näher am Wert Null als im Vergleich zu den Fahrten ohne Gerätenutzung und sie liegen bei positiven Werten, was auf eine Zunahme der Belastung beim Gerät des Subtrahenden im Vergleich zum Gerät des Minuenden hindeutet: 0,6 (HMD2 – HMD1), 1,7 (HMD2 – Tablet) und 1,1 (HMD1 – Tablet).

Die drei Stichproben aus Abbildung 14 erfüllen zwar den Shapiro-Wilk-Test mit einem Signifikanzwert deutlich über 0,05 ($p = 0,25$ HMD2 – HMD1, $p = 0,34$ HMD2 – Tablet, $p = 0,84$ HMD1 – Tablet), aber die folgenden t-Tests zeigen keinen signifikanten Unterschied zum Wert Null bei einem korrigierten Signifikanzniveau $\alpha^* = 0,0083$. Die Differenzen der TLX-Mittelwerte zwischen der Nutzung von HMD2 und dem Tablet liegen bei $p = 0,03 < 0,5$, sodass bei diesem Vergleich am ehesten zu erwarten wäre, dass ein Unterschied vorliegt. Da es aber an weiteren Hinweisen fehlt, kann an dieser Stelle nur festgestellt werden, dass die Analyse der TLX-Mittelwerte keinen Unterschied in der Belastung bei der Nutzung der drei Geräte feststellen kann.

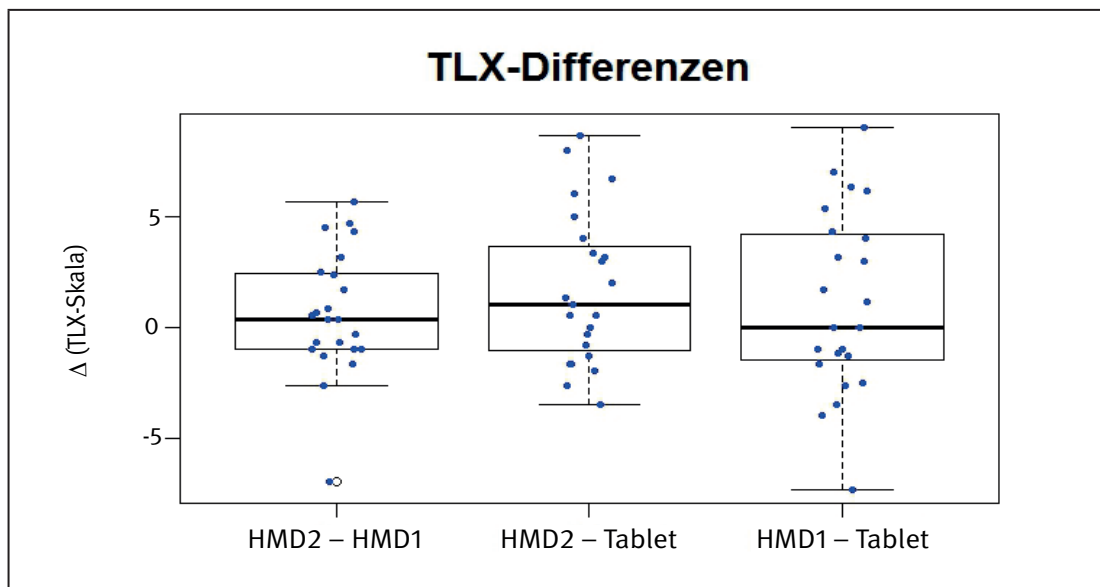


Abbildung 14: Verteilung der Differenzen der TLX-Mittelwerte pro Proband als Punktwolke (blaue Kreise) und Boxplot (zwischen Geräten). Die Differenzen beziehen sich auf Fahrten mit unterschiedlichen Geräten. Minuend und Subtrahend entsprechen der Beschriftung.

4.4.3 Ergebnisse – NASA-TLX – Einzelfragen

In diesem Abschnitt wird untersucht, welche Rückschlüsse sich daraus ziehen lassen, wenn man die Antworten auf die Einzelfragen des NASA-TLX untersucht. Die sechs Fragen des NASA-TLX teilen die Gesamtbelastung in unterschiedliche Aspekte auf:

1. mentale Belastung
2. körperliche Belastung
3. zeitliche Belastung
4. Leistung
5. Anforderung
6. Frustration

Dabei ist in unserem Versuch zu erwarten, dass die kognitive und sensorische Belastung des Versuchsauf-

baus sich nur in den Belastungsaspekten 1 und 3 bemerkbar macht. Die geringe körperliche Belastung ist bei allen Fahrten gleich und die Belastungsaspekte vier bis sechs ergeben sich aus dem Zusammenwirken mit anderen Personen, die nur eingeschränkt vorgelegen haben (z. B. durch die Versuchsleitung).

Zunächst werden für jeden Probanden die Differenzen der einzelnen Antworten bei verschiedenen Fahrten gebildet. Dabei werden im oberen Teil von Abbildung 15 die Werte für die Fahrten mit Gerätenutzung von den Werten für die Fahrt ohne Geräte abgezogen. Im unteren Teil sind Differenzen für die Gerätenutzung untereinander abgebildet. Zusammen mit den Mittelwerten sind auch die Standardabweichungen dieser Differenzen aufgetragen.

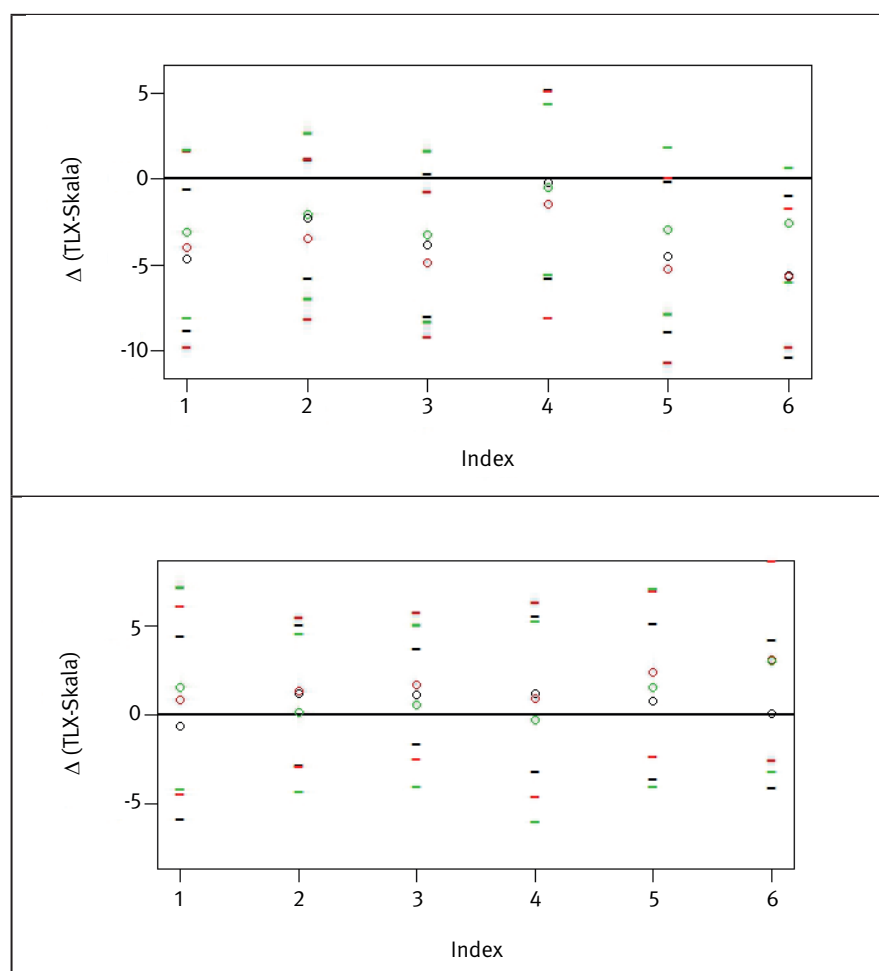


Abbildung 15:
NASA-TLX Belastungsdifferenzen der einzelnen Fragen; Mittelwerte (Kreise) und Standardabweichungen (Striche) der Differenzen der Probandenangaben für die sechs Fragen im NASA-TLX. Minuend und Subtrahend entsprechen den Beschriftungen:

Baseline – HMD1 (schwarz), Baseline – HMD2 (rot), Baseline – Tablet (grün)

HMD2 – HMD1 (schwarz), HMD2 – Tablet (rot), HMD1 – Tablet (grün)

Dabei ist zu sehen, dass nur in der oberen Abbildung bei fünf Fragen die Mittelwerte deutlich von Null verschieden sind. Nur beim Kriterium Leistung bleibt auch in der oberen Abbildung die Belastung etwa gleich. Die Mittelwerte bei den anderen fünf Aspekten sind negativ, was eine Belastungszunahme bei der Benutzung von Anzeigegeräten bedeutet.

An dieser Stelle zeigt sich, dass die Angaben im NASA-TLX nicht spezifisch für die kognitive und sensorische Belastung sind, die im Versuch nach ISO 17488 untersucht wird. Es ist daher nicht sinnvoll, die Fragen zu gewichten, wie es die Auswertung des NASA-TLX eigentlich vorsieht [8]. Stattdessen wird als Belastungsgröße pro Proband und Fahrt der Mittelwert aus den Angaben zu den sechs Fragen gebildet. Dieses Vorgehen, das die Belastungsaspekte nicht unterschiedlich gewichtet, wird in der Lite-

ratur als „Raw NASA-TLX“ bezeichnet und führt dazu, dass mit dem NASA-TLX in dieser Studie nur eine allgemeine Belastungseinschätzung der Probanden analysiert wird.

Da das Kriterium Leistung in keinem Vergleich zu einer veränderten Belastung geführt hat, könnte ein Vorschlag für eine mögliche Gewichtung sein, auf das Kriterium Leistung zu verzichten. Dies verändert die Mittelwerte aller Differenzen M für einen Vergleich, wie sie in Abschnitten 4.4.1 und 4.4.2 berechnet wurden, aber nur geringfügig. Zwar wird der Mittelwert nur noch mit fünf anstatt sechs Fragen gebildet, aber der Beitrag des Kriteriums Leistung ($j = 4$) zur Gesamtdifferenz entspricht dem Mittelwert der Einzelfrage und damit etwa dem Wert Null:

$$M = \frac{1}{23 \cdot 6} \sum_i \sum_j A_{ij} - B_{ij} = \frac{1}{6} \sum_i \underbrace{\frac{1}{23} \sum_j A_{ij} - B_{ij}}_{\approx 0 \text{ für } j = 4} \quad (1)$$

In Gl. (1) sind A_{ij} und B_{ij} TLX-Mittelwerte von zwei unterschiedlichen Arten von Fahrten, die bei i Probanden und j Fragen erhoben wurden. Entfernt man die Frage nach dem Leistungskriterium aus der Auswertung, treten die Effekte, die in den beiden vorgehenden Abschnitten beschrieben wurden, stärker zutage. Es gibt aber keine qualitativ andere Einschätzung. So steigen die Belastungen im Vergleich mit den Fahrten ohne Geräte zwischen 0,4 und 0,7 Punkte und die p -Werte werden kleiner. Bei den Vergleichen der Geräte untereinander, bei denen keiner der Aspekte deutliche Beiträge liefert, fallen die Unterschiede noch geringer aus.

4.5 Ergebnisse – Masteinblendung

Wie im Abschnitt 3.1 „Messung der Reaktionszeiten“ beschrieben, wurde fünf Probanden während der Fahrten ein Mast eingeblendet, 18 Probanden sind ohne eingeblendeten Mast gefahren.

Um zu klären, ob dieses Einblenden einen Effekt auf die Kennzahlen gehabt hat, sind in Abbildung 16 die Verteilungen der mittleren Reaktionszeiten der 18 Probanden, die keinen Mast gesehen haben, als Boxplots für die vier verschiedenen Fahrten dargestellt. Zusätzlich sind die Mittelwerte der Probanden als Punktwolke dargestellt, die einen Mast eingeblendet hatten. Zusätzlich enthält Abbildung 17 eine analoge Darstellung für die NASA-TLX-Mittelwerte.

In beiden Abbildungen ist zu sehen, dass die Werte für die Fahrten mit Mast in etwa in denselben Bereichen verteilt sind wie die Werte für Fahrten ohne Mast. Rechnet man für alle acht Verteilungen der beiden Abbildungen 95-Perzent-Konfidenzintervalle für die Mittelwerte aus (t-Verteilungen, mit 17 Freiheitsgraden), so liegen die Mittelwerte für die Stichprobe mit Mast nur in zwei Fällen außerhalb dieser Intervalle (Reaktionszeiten: Tablet, NASA-TLX-Mittelwerte: HMD2).

Zwar sind die Voraussetzungen für einen Zweistichproben-t-Test (Welch-Test) streng genommen nicht gegeben, weil die Stichprobengrößen zu klein und zu unterschiedlich sind. Führt man den Test aber dennoch durch, ergeben sich keine signifikanten Unterschiede zwischen den Gruppen. Insgesamt gibt es keine hinreichenden Gründe, die beiden Gruppen getrennt zu behandeln, daher wurden sie für die Auswertung zusammengefasst (siehe Kapitel 5 „Diskussion“).

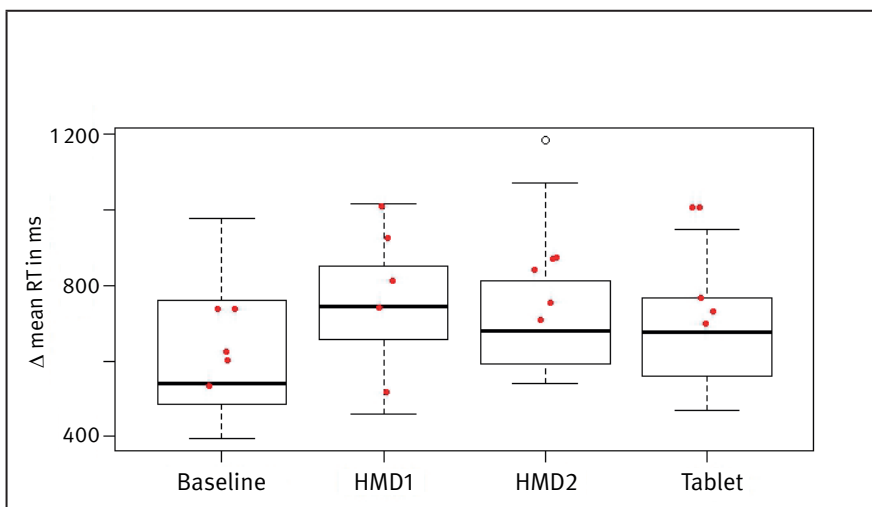


Abbildung 16: Verteilungen der mittleren Reaktionszeiten für verschiedene Arten von Fahrten als Boxplots für 18 Probanden ohne Masteinblendung, als Punktwolke (rote Kreise) für fünf Probanden mit Masteinblendung

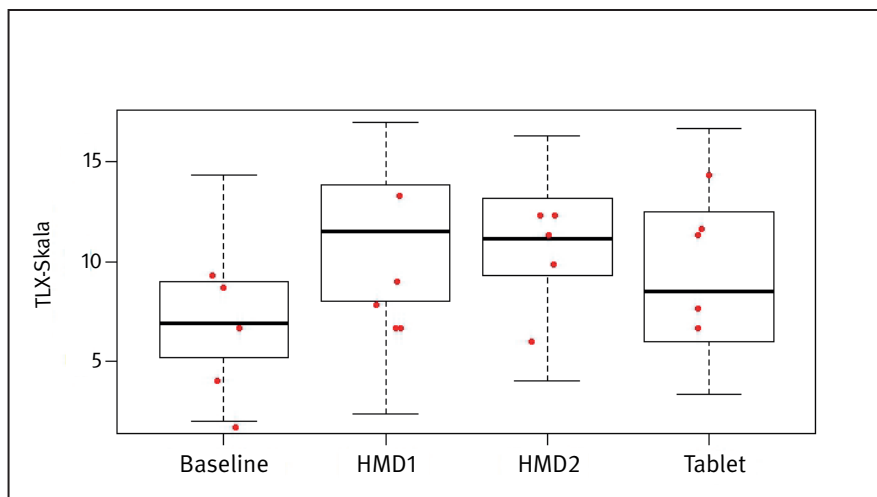


Abbildung 17:
Verteilungen der NASA-TLX-Mittelwerte für verschiedene Arten von Fahrten als Boxplots für 18 Probanden ohne Masteinblendung, als Punktwolke (rote Kreise) für fünf Probanden mit Masteinblendung

4.6 Ergebnisse – Allgemeiner Fragebogen

Die Ergebnisse zu den allgemeinen Fragen sind in Tabelle 3 zusammengefasst. Dabei ist festzuhalten, dass den meisten Probanden das Tablet am besten gefallen hat, gefolgt mit deutlichem Abstand vom monokularem Anzeigegerät und nur zwei Probanden hat das binokulare Anzeigegerät am besten gefallen.

Diese Reihenfolge ergibt sich auch bei der Frage nach der geringsten Ablenkung. In Klammern angegeben ist die

Anzahl, die dasselbe Gerät als bestes bezeichnet hat und als das mit der geringsten Ablenkung. Hier haben nur zwei Probanden eine andere Antwort gegeben als bei der Frage nach dem besten Gerät. Ein Proband hat nicht geantwortet.

Die Frage nach Sichtbehinderungen enthält dieselbe Information wie die Frage nach der geringsten Ablenkung: Das Tablet wird mit der geringsten Ablenkung und Sichtbehinderung verbunden, das binokulare Anzeigegerät mit der größten.

Tabelle 3:

Ergebnisse zu den allgemeinen Fragen; die Antwort zu mehreren Geräten ist nur bei der Frage nach der Sichtbehinderung möglich: bestes Gerät – Anzahl, denen das jeweilige Gerät am besten gefallen hat, geringste Ablenkung – Anzahl mit der subjektiv geringsten Ablenkung beim jeweiligen Gerät, in Klammern die Anzahl, denen auch das Gerät am besten gefallen hat, Sichtbehinderung bei einem Gerät – erlebte Sichtbehinderung

	Bestes Gerät	Geringste Ablenkung	Sichtbehinderung
Tablet	14	13 (12)	2
HMD1	7	7 (6)	4
HMD2	2	2 (2)	11
Kein Gerät/Mehrere Geräte	-	-	1/5

5 Diskussion

5.1 Studienkollektiv

Das Studienkollektiv war auf berufstätige Männer der Fa. CINOVIATION und des IFA beschränkt, die keine Brille tragen. Außerdem wurde das Studienkollektiv nicht nach Zufallskriterien ausgewählt. Daher kann es auch nicht als eine repräsentative Stichprobe für eine andere Gruppe angesehen werden (etwa alle Gabelstaplerfahrer in Deutschland). Die hier vorgestellten Ergebnisse sind somit nur eingeschränkt auf andere Gruppen übertragbar.

Eine weitere Einschränkung könnte sich auch aus der vergleichsweise geringen Benutzungsdauer von Gabelstaplern pro Monat ergeben. Diese kann darauf zurückgeführt werden, dass das IFA kein Unternehmen der Logistikbranche ist, sodass Fahren mit dem Gabelstapler nicht zur Hauptaufgabe der Probanden gehören, und dass auch bei der Fa. CINOVIATION während der Studiendauer keine üblichen Betriebsbedingungen geherrscht haben. Dass die Probanden die Primäraufgabe gut erfüllt haben, ergibt sich auch aus der folgenden Diskussion der Fahrfehler.

5.2 Fahrfehler

Weder die geringen angegebenen Benutzungsdauern noch die Unterschiede zwischen Simulator und realen Fahrzeugen (z. B. beim Lenkungsspiel) haben zu häufigen Fahrfehlern geführt (Abbildung 5). Dies ist ein Beleg dafür, dass der erstmalige Einsatz dieses Fahrsimulators in Verbindung mit einem Test nach ISO 17488 erfolgreich war: Die vorgeschriebenen Übungsphasen waren für Personen mit Staplerschein ausreichend, um die Primäraufgabe des Tests (Fahren) erfolgreich zu erfüllen. Darüber hinaus belegt die geringe Zahl an Fahrfehlern, dass die Probanden, wie verlangt, die Primäraufgabe nicht vernachlässigt haben, um zu besseren Reaktionszeitergebnissen zu gelangen.

Wie in Abschnitt 4.2 „Ergebnisse – Fahrfehler“ beschrieben, kann das Auftreten von Fahrfehlern in dieser Studie nicht unterschieden werden von einem rein zufälligen Effekt, der unabhängig von den Probanden und deren Belastungen ist. Das bedeutet aber nicht, dass bei einem anderen Versuchsaufbau (Gegenverkehr, längere Fahrzeiten ...) nicht doch ein Zusammenhang von Fahrfehlern und Belastung untersucht werden kann.

5.3 Masteinblendung

Mit der Masteinblendung wurde ein weiterer Effekt untersucht, der vor allem den Versuchsaufbau betrifft. Zwar sind die Stichproben nicht groß genug für einen schließenden Test, die deskriptive Untersuchung hat aber keine starken Hinweise dafür gefunden, dass die Masteinblendung

einen Einfluss auf die Reaktionszeiten oder TLX-Mittelwerte gehabt hat.

Nur bei den mittleren Reaktionszeiten deutet sich an, dass diese für die Fahrten mit Mast überwiegend oberhalb der Mediane für die Fahrten ohne Mast liegen (Abbildung 16). Bei den TLX-Mittelwerten ist dieser Trend nicht zu sehen. Diese höheren Reaktionszeiten könnten mit einer höheren kognitiven und sensorischen Belastung erklärt werden, wenn der Mast das Blickfeld abdeckt und somit für die Primäraufgabe weniger Information zur Verfügung steht. Da dies sowohl die Vergleichsaufgabe als auch die Fahrten mit Sekundäraufgabe betrifft, war auch in den Differenzen der Reaktionszeiten für die Probanden kein eindeutiger Effekt zu sehen (auf einen Beweis wird hier verzichtet).

Da die Datengrundlage zu gering ist, um weitere Aussagen zu treffen, wurden beide Stichproben in das Studienkollektiv aufgenommen. Zukünftige Untersuchungen sollten sich konsistent für eine der beiden Einblendungen entscheiden.

5.4 Reaktionszeiten

Die Reaktionszeiten sind die wesentliche Kennzahl, auf die sich Aussagen nach ISO 17488 stützen. Sie sind objektiv messbar, messen eine spezifische Belastung und sind überprüfbar. Daher ist mit dieser Kennzahl die größte Aussagekraft verbunden.

Im Ergebnis hat die kognitive und sensorische Belastung der Fahrer dieses Kollektivs durch das Nutzen der Anzeigeräte (HMD1, HMD2, Tablet) zugenommen und diese Zunahme war statistisch signifikant. Dies beantwortet eine wesentliche Frage dieser Untersuchung.

Ob mit dieser Belastung eine Beanspruchung verbunden ist und welche das ist, muss in folgenden Studien geklärt werden. Denkbar sind zum Beispiel Muskel-Skelett-Beschwerden (Verspannungen im Nackenbereich ...), die bei längerer Exposition auch chronisch werden könnten. Ein mögliches höheres Unfallrisiko kann ebenfalls in Betracht gezogen werden.

Ob es einen Unterschied zwischen den Anzeigeräten gibt, konnte in dieser Studie nicht geklärt werden. Weder bei den Reaktionszeiten noch bei den Trefferraten gibt es schlüssige Effekte, die belegen, dass zum Beispiel ein Gerät immer besser wäre als andere. Auch sind diese Effekte nicht signifikant. Daher muss die Frage nach den Unterschieden zwischen den Geräten von nachfolgenden Studien geklärt werden.

5.5 NASA-TLX

Der NASA-TLX-Fragebogen wurde in dieser Studie so beantwortet, dass er nicht eindeutig einer spezifischen Belastung zugeordnet werden kann und eine Gewichtung der einzelnen Belastungsaspekte in der Analyse nicht möglich ist. Diese Unschärfe kann auf die subjektive Auslegung der Fragen durch die Probanden zurückgeführt werden. Es ist aber plausibel, dass die Analyse, die sich auf Mittelwerte aller Fragen stützt, dennoch stark mit der kognitiven und sensorischen Belastung des Versuches zu tun hat, zumal insgesamt die TLX-Mittelwerte und die Reaktionszeiten zu qualitativ denselben Einschätzungen kommen.

Es kommen noch weitere Aspekte hinzu, die nicht einfach erklärt werden können. So gehen die TLX-Mittelwerte zu den einzelnen Fragen entweder in eine Richtung oder sie sind Null und spielen deshalb für die Auswertung keine Rolle. Die geringen, nicht signifikanten Belastungsunterschiede zwischen den Geräten können auf Einzelfragen zurückgeführt werden, die im Mittel ebenfalls um Null liegen (Abbildung 15 unten). Es kommt also nicht zu Kompensationen von positiven und negativen Effekten einzelner Aspekte. Daher wurde der Gesamtmittelwert („raw NASA-TLX“) für diese Studie als Kennzahl verwendet.

Trotz der geringeren Aussagekraft und höheren Unschärfe des NASA-TLX im Vergleich zu den objektiven Messungen der Reaktionszeit liefert die Analyse hier qualitativ dieselben Ergebnisse: Die Belastung nimmt signifikant mit der Gerätenutzung zu, aber zwischen den Geräten kann kein Belastungsunterschied festgestellt werden.

5.6 Allgemeine Fragen

Die am wenigsten belastbare Größe stellen die allgemeinen Fragen dar. Die Frage nach dem Gerät, das am besten gefallen hat, kann mehr mit dem Aussehen des Gerätes oder der Vorliebe der Probanden für Technik zu tun haben als mit der objektiven Belastung, die von diesen Geräten ausgeht. Auch die starke Verknüpfung zwischen den Fragen nach dem besten Gerät und der Ablenkung kann zum Teil dadurch hervorgerufen sein, dass die Fragen kurz hintereinander gestellt wurden. Dennoch erscheint es nicht unplausibel, dass die Vorliebe für das Tablet auch damit zusammenhängt, dass die Probanden sich am wenigsten abgelenkt gefühlt haben.

Obwohl es keine objektiven Unterschiede in der kognitiven und sensorischen Belastung (Reaktionszeiten) oder der allgemein empfundenen Belastung (NASA-TLX) gab, ist wahrscheinlich anzunehmen, dass die Akzeptanz für das Tablet höher ist als für die kopfgetragenen Anzeigegeräte, auch wenn keine Gründe dafür angegeben werden können.

6 Schlussfolgerung

Zunächst ist festzustellen, dass mit dem gewählten Versuchsaufbau (Fahrsimulator) und dem Studiendesign erstmals ein erfolgreicher Reaktionszeittest nach ISO 17488 mit Bezug zu Datenbrillen im innerbetrieblichen Transport stattgefunden hat. In zukünftigen Studien kann auf die Ergebnisse zu Probandenauswahl, Masteinblendung und erwartbaren Fahrfehlern zurückgegriffen werden.

Was die zentrale Studienfrage betrifft, so wurde festgestellt, dass die Belastung der Fahrer bei der Nutzung von Anzeigegeräten zunimmt und diese Zunahme statistisch signifikant ist. Das betrifft sowohl die objektiv messbare kognitive und sensorische Belastung nach ISO 17488 als auch die allgemein empfundene Belastung nach dem NASA-TLX-Fragebogen. Allerdings ist keine unterschiedliche Belastung beim Nutzen der drei Anzeigegeräte feststellbar.

Diese Ergebnisse sind nur eingeschränkt auf andere Fahrer von Flurförderzeugen übertragbar und es können keine Aussagen dazu getroffen werden, ob diese Belastungen auch mit Beanspruchungen im betrieblichen Einsatz verbunden sind (Unfälle, Muskel-Skelett-Beschwerden). Dies müsste in weiteren Studien geklärt werden.

Zum jetzigen Zeitpunkt kann nur gesagt werden, dass mit den Anzeigegeräten eine Belastung verbunden ist. Auch wenn nicht geklärt ist, welche Beanspruchungen mit ihr verbunden ist und sie sich deshalb auch nur schwer mit anderen Belastungen vergleichen lässt, sollte vor dem Einsatz solcher Anzeigegeräte im Rahmen einer Gefährdungsbeurteilung abgewogen werden, ob an dem speziellen Arbeitsplatz eine zusätzliche Belastung vertretbar ist. In dieser Studie wurde das Anzeigegerät durchgängig genutzt, um Informationen auszulesen. Im betrieblichen Einsatz ist daher die Belastungszunahme stark davon abhängig, wie oft Information aus dem Gerät ausgelesen werden muss.

Wenn der Einsatz erfolgen soll, ist nach dieser Studie kein Unterschied zwischen den untersuchten Anzeigegeräten festzustellen. Allerdings hat das Tablet den Vorteil, dass es auch von Brillenträgern genutzt werden kann und die Akzeptanz für dieses Gerät wahrscheinlich größer ist als für kopfgetragene Anzeigen.

7 Literatur

- [1] *Forni, A.*: Transform business outcomes with immersive technology. <https://www.gartner.com/smarterwithgartner/transform-business-outcomes-with-immersive-technology> (abgerufen am 07.05.2018)
- [2] *Terhoeven, J.; Wischniewski, S.*: Datenbrillen im Einsatz. *Gute Arbeit* (2017) Nr 5, S. 24-26
- [3] Statistik Arbeitsunfallgeschehen 2016. Hrsg.: Deutsche Gesetzliche Unfallversicherung e. V. (DGUV), Berlin 2017
- [4] *Schömig, N.; Schoch, S.; Neukum, A.; Schumacher, S.; Wandtner, B.*: Simulatorstudien zur Ablenkungswirkung fahrfremder Tätigkeiten. Hrsg.: Bundesanstalt für Straßenwesen (BASt), Bergisch Gladbach. NW-Verlag, Bremen 2015
- [5] *Kirchhoff, B.; Wischniewski, S.; Adolph, L.*: Head Mounted Displays – Arbeitshilfen der Zukunft. Bedingungen für den sicheren und ergonomischen Einsatz monokularer Systeme. Hrsg.: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Dortmund 2016
- [6] Ausbildung und Beauftragung der Fahrer von Flurförderzeugen mit Fahrersitz und Fahrerstand (DGUV Grundsatz 308-001, bisher BGG 925). Hrsg.: Deutsche Gesetzliche Unfallversicherung e. V. (DGUV), Berlin 2007
- [7] ISO 17488: Road vehicles – Transport information and control systems – Detection-response task (DRT) for assessing attentional effects of cognitive load in driving (2016)
- [8] *Hart, S. G.; Staveland, L. E.*: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: *Hancock, P. A.; Meshkati, N.* (Hrsg.): Human mental workload. Elsevier, Amsterdam 1988, S. 139-183
- [9] *Diez, D. M.; Barr, C. D.; Çetinkaya-Rundel, M.*: OpenIntro Statistics. <https://www.openintro.org/>
- [10] *Sachs, L.*: Angewandte Statistik. 9. Aufl. Springer, Berlin 1999
- [11] *Shapiro, S. S.; Wilk, M. B.*: An analysis of variance test for normality (for complete samples). *Biometrika* 52 (1965) Nr. 3/4, S. 591-611

Anhang 1: Probandeninformation und Einverständniserklärung

IFA – Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung
Alte Heerstraße 111, 53757 Sankt Augustin

Studieninformation für die Fälle

Aufgabenlasten durch Anzeigesysteme auf Gabelstaplern

Sehr geehrte Damen und Herren,

wir möchten Sie herzlich um Ihre Teilnahme an einer wissenschaftlichen Studie bitten.

Zurzeit wird eine Studie zur Prüfung von *Ablenkungen* am Arbeitsplatz (hier *Gabelstapler*) durchgeführt. Diese Studie beschäftigt sich mit Anzeigesystemen (Datenbrillen und Tablet-PC), die typischerweise im Arbeitsalltag genutzt werden. Bisher ist weitestgehend ungeklärt mit welchen Konsequenzen. Diese Studie wird daher von den Experten des Instituts für Arbeitsschutz durchgeführt. Wir ermitteln das Ausmaß der so genannten Aufgabenlasten und möglichen Ablenkungspotenzialen. Unser Ziel ist ein sicherer und gesundheitsgerechter Umgang mit den benannten Systemen am Arbeitsplatz.

Worin besteht Ihr Beitrag zur Studie?

- Wir führen mit Ihnen einen etwa einstündigen Versuch am Fahrsimulator durch.
- Wir testen nicht Sie bzw. Ihre *Leistung*, sondern die Technik. Sie können nichts *falsch machen* im Rahmen des Versuchs.
- Sie erhalten von uns kurze Fragebögen zur Einschätzung der technischen Systeme.

Was geschieht mit Ihren Angaben?

- Ihre Angaben werden in eine Datenbank eingegeben und für anonyme statistische Auswertungen genutzt.
- Informationen zur Personenidentifizierung (Name, Kontaktinformation) werden von der Studie nicht berührt.

Wann sollten Sie von einer Teilnahme absehen?

- Wenn Sie ohne eine Sehhilfe nicht fahrtüchtig sind.
- Wenn Sie unter Epilepsie leiden.
- Wenn Sie aus Erfahrung wissen, dass es Ihnen durch Computerspiele schwindelig wird (*simulator sickness*).

Ihre Teilnahme an der Studie ist absolut freiwillig. Wenn Sie nicht teilnehmen möchten, entstehen Ihnen daraus keine Nachteile.

Die Vorschriften des Datenschutzes werden selbstverständlich eingehalten. Ergebnisse der Studie werden in Form von Statistiken zusammengestellt, die keinerlei Rückschlüsse auf Einzelpersonen zulassen.

Wir möchten Sie nochmals herzlich um Ihre Teilnahme an dieser Studie bitten. Der Erfolg hängt ganz wesentlich von einer Einwilligung möglichst aller zur Teilnahme eingeladenen Personen ab.

IFA – Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung
Alte Heerstraße 111, 53757 Sankt Augustin

Einwilligungserklärung für die Teilnahme der Studie (Fallrekrutierung)

Ich bin damit einverstanden, mich an der Studie zum Thema „*Aufgabenlasten durch Anzeigesysteme auf Gabelstaplern*“ zu beteiligen. Die schriftlichen Informationen zur Zielsetzung, Art und Umfang der Studie habe ich erhalten und verstanden. Ich werde vor Ort über die Studie weiterführend informiert – mir ist bekannt dort Fragen aller Art stellen zu können.

Meine Teilnahme ist freiwillig. Ich kann mein Einverständnis zur Teilnahme jederzeit ohne Angabe von Gründen – und ohne dass mir daraus irgendwelche Nachteile entstehen – zurückziehen.

- Ich erkläre mich dazu bereit, an dem Versuch im Fahrsimulator (Frontstapler) teilzunehmen.
- Ich bin damit einverstanden, dass mir im Rahmen des Versuchs ein Fragebogen bzgl. der technischen Systeme (Nasa TLX) vorgelegt wird.

Mir ist bekannt, dass meine Angaben ausschließlich für wissenschaftliche Zwecke von den Experten des Instituts für Arbeitsschutz verwendet werden. Für die Anonymisierung der Daten, wird mein Name nicht erfasst. Die Speicherung und Auswertung der Daten erfolgt unter voller Wahrung des Datenschutzes.

.....

Ort, Datum

.....

Unterschrift

Name, Vorname

Als Ansprechpartner stehen die Studienmitarbeiter zur Verfügung.

Name: Dr. Michael Bretschneider-Hagemes, Benno Gross, Andreas Stefan

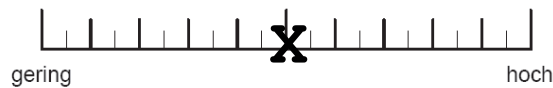
Rufnummer: 02241-231-2902

Anhang 2: Fragebögen

Beanspruchungshöhe

Geben Sie jetzt für jede der unten stehenden Dimensionen an, wie hoch die Beanspruchung war. Markieren Sie dazu bitte auf den folgenden Skalen, in welchem Maße Sie sich in den sechs genannten Dimensionen von der Aufgabe beansprucht oder gefordert gesehen haben:

Beispiel:



Geistige Anforderungen

Wie viel geistige Anstrengung war bei der Informationsaufnahme und -verarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erforderte sie hohe Genauigkeit oder war sie fehlertolerant?



Körperliche Anforderungen

Wie viel körperliche Aktivität war erforderlich (z.B. Ziehen, Drücken, Drehen, Steuern, Aktivieren,...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholsam oder mühselig?



Zeitliche Anforderungen

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt, mit dem Aufgaben oder Aufgabenelemente auftraten? War die Abfolge langsam und geruhsam oder schnell und hektisch?



NASA-TLX (Kurzfassung deutsch)

Leistung

Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

**Anstrengung**

Wie hart mussten sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

**Frustration**

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?



Kontrollieren sie bitte, ob Sie zu allen Fragen Angaben gemacht haben. Bei Unklarheiten wenden Sie sich bitte an die Versuchsleiterin / den Versuchsleiter.

Subskalen:

Der Wert jeder Subskala ist ein Einzelmesswert. Je nach Fragestellung können Subskalen auch einzeln verwendet oder untereinander kombiniert werden.

Auswertung:

Jedem Kreuz wird ein ganzzahliger Wert von 0 bis 20 zugeordnet, wobei „0“ als gering und „20“ als hoch gewertet wird. Befindet sich ein Kreuz zwischen zwei Teilstrichen der Skala, wird derjenige ganzzahlige Wert vergeben, der sich näher am Kreuzungspunkt der Angabe des Teilnehmers befindet.

Quellen:

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, 904-908. Santa Monica: HFES.

Vor den Fahrten

- Probandennummer: _____
- *Standardtest* oder *Stichprobe* (bitte ankreuzen)
- Datum: _____
- Ort: _____
- Uhrzeit: _____
- Geschlecht: m w
- Alter: _____
- Brillenträger?: J N
- Staplerschein?: J N
- Wöchentliche Fahrpraxis (ca. Stunden/Monat auf dem Stapler)?: _____ h

Nach jeder Fahrt

- Proband übel geworden? J N
- Proband liest Zahlen vor?: J N

Fragen zum Abschluss

Welches Anzeigesystem hat Ihnen insgesamt am besten gefallen?

Datenbrille „monokular“

Datenbrille „binokular“

Tablet-PC/Monitor

Welches Anzeigesystem hat Sie am wenigsten von der Fahrt abgelenkt?

Datenbrille „monokular“

Datenbrille „binokular“

Tablet-PC/Monitor

Haben Sie eines der Anzeigesysteme als Sichtbehinderung erlebt?Datenbrille „monokular“ Datenbrille „binokular“ Tablet-PC/Monitor Nein **Sonstige Hinweise die Sie uns mitteilen möchten?**
