

Entwicklung und Validierung eines Methoden- Inventars zur Human-Factors-Evaluation von Flugzeug-Cockpit-Gestaltungen

Inaugural-Dissertation zur Erlangung der Doktorwürde der Fakultät für Humanwissen-
schaft, der Universität Regensburg

Vorgelegt von

Constanze Billeriß
Landshut, August 2020

Erstgutachter: Prof. Dr. Dr. h.c. Alf C. Zimmer

Zweitgutachter: Prof. Dr. rer. nat. Mark W. Greenlee

Dank

Mein Dank gilt allen, die mich bei der Anfertigung dieser Arbeit unterstützt haben. Dies gilt insbesondere für Herrn Prof. Dr. Dr. h.c. Alf Zimmer und Dr. Peter Sandl, die diese Arbeit ermöglicht, bereichert und betreut haben. Weiterhin möchte ich Dr. Alexander Rabl und Dr. Johannes Kellner für Rat und Ermutigungen danken, Christian Hienstorfer für seine Unterstützung zur Planung und Umsetzung der Evaluierungen, dem HFE-Team für Kreativität und Unterstützung sowie den Piloten für ihre Geduld und Bereitschaft zur Teilnahme. Außerdem möchte ich Patrick Langheim für seine Python-Kenntnisse und Dr. Franziska Schmidl sowie Dr. Christoph Vernaleken für Review und Feedback danken. Meinem Mann Sven, Systemadministrator in allen Lebenslagen, und meinem Sohn danke ich für ihre Geduld und ihr Verständnis und den Omas für ihre Bereitschaft, mir die Freiräume zum Arbeiten zu ermöglichen.

INHALTSVERZEICHNIS

Abstrakt	1
Abstract	2
1 Einleitung.....	3
1.1 Problemstellung	3
1.2 Ziele	4
1.3 Aufbau der Arbeit.....	5
2 Erhebung und Bewertung der MMS Cockpit in einem komplexen sozio-technischen System	7
2.1 Anpassungen der MMS Flugzeug-Cockpit.....	7
2.2 Der Nutzer und dessen Nutzungskontext	8
2.3 Die Überprüfung von Cockpit-Modifikationen in einem Cockpit-Simulator	10
2.4 Bewertungskriterien und deren Erhebung im Cockpit-Simulator.....	11
3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit.....	14
3.1 Situationsbewusstsein.....	14
3.2 Workload - Beanspruchung und Arbeitsbelastung	17
3.3 Gebrauchstauglichkeit.....	19
3.4 Das multidimensionale Erhebungskonzept zur Bewertung von MMS.....	21
3.4.1 Methodenauswahl zur Erfassung von Situationsbewusstsein	22
3.4.2 Methodenauswahl zur Erfassung von Workload	27
3.4.3 Methodenauswahl zur Erfassung von Gebrauchstauglichkeit	34
3.4.4 Das multidimensionale Methoden-Inventar	41
4 Integration des Methoden-Inventars in die Simulationsumgebung	43
4.1 Experimentelle Erhebung in der Cockpit-Simulation	44
4.2 Entwicklung von Messbedingungen im situativen Kontext	47
4.3 Umsetzung des Methoden-Inventars in der Simulationsumgebung	50
5 Die Evaluierung des multidimensionalen Methoden-Inventars.....	52
5.1 Machbarkeitsstudie zur Implementierung einer Navigationshilfe.....	53
5.1.1 Stichprobenbeschreibung	53
5.1.2 Verwendete Methoden	53
5.1.3 Szenario-Beschreibung	54
5.1.4 Versuchsablauf	54
5.1.5 Bewertung der Bedienung einer Navigationshilfe in einer Machbarkeitsstudie	54

Inhaltsverzeichnis

5.1.6	Bewertung der verwendeten Methoden-Kombination	55
5.2	Untersuchung des Re-Designs einer Navigationshilfe in verschiedenen Situationen.....	55
5.2.1	Stichprobenbeschreibung	56
5.2.2	Verwendete Methoden	56
5.2.3	Szenario-Beschreibung	58
5.2.4	Versuchsablauf	58
5.2.5	Bewertung der Verwendung des Re-Designs einer Navigationshilfe.....	59
5.2.6	Bewertung der verwendeten Methoden-Kombination	80
5.3	Untersuchung eines räumlich-auditiven Warnsignals	81
5.3.1	Stichprobenbeschreibung	82
5.3.2	Verwendete Methoden	82
5.3.3	Szenario-Beschreibung	83
5.3.4	Versuchsablauf	84
5.3.5	Bewertung der Verwendung eines räumlich-auditiven Warnsignals.....	84
5.3.6	Bewertung der verwendeten Methoden-Kombination	84
5.4	Untersuchung zweier Bedienformen zur Einstellung eines Sensors	85
5.4.1	Stichprobenbeschreibung	85
5.4.2	Verwendete Methoden	85
5.4.3	Szenario-Beschreibung	86
5.4.4	Versuchsablauf	86
5.4.5	Bewertung der Bedienformen zur Einstellung eines Sensors	86
5.4.6	Bewertung der verwendeten Methoden-Kombination	87
6	Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität	88
6.1	Die Konstrukt-Validität als Grundlage einer Validierungsstrategie.....	88
6.2	Der Multi-Trait/Multi-Method-Ansatz als Validierungsstrategie	90
6.2.1	Sub-Matrizen zur Bestimmung konvergenter Validität.....	94
6.2.2	Sub-Matrizen zur Bestimmung diskriminanter Validität.....	96
6.2.3	Interpretation der Analysematrizen.....	97
7	Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene	99
7.1	Validierungsergebnisse der Qualitätskriterien auf Konstrukt-Ebene.....	99
7.1.1	Situationsbewusstsein	99
7.1.2	Workload	102
7.1.3	Gebrauchstauglichkeit.....	106
7.1.4	Zusammenfassende Betrachtung der Konzepte	109
7.2	Validierungsergebnisse der Einzelmethoden	110

Inhaltsverzeichnis

7.2.1	Methoden zur Erfassung des Situationsbewusstseins	111
7.2.2	Methoden zur Erfassung der Workload	120
7.2.3	Methoden zur Erfassung der Gebrauchstauglichkeit.....	130
7.2.4	Zusammenfassende Ergebnis-Betrachtung der verwendeten Einzelmethoden.....	143
8	Bewertung der Qualitätskriterien und des multidimensionalen Methoden-Inventars	144
8.1	Analyse und Bewertung der Qualitätskriterien auf Konstrukt-Ebene.....	144
8.1.1	Situationsbewusstsein.....	144
8.1.2	Workload	146
8.1.3	Gebrauchstauglichkeit.....	147
8.1.4	Zusammenfassende Betrachtung der Konzepte	148
8.2	Analyse und Bewertung der Einzelmethoden und des Methoden-Inventars.....	149
8.2.1	Methoden zur Erfassung des Situationsbewusstseins	150
8.2.2	Methoden zur Erfassung der Workload	153
8.2.3	Methoden zur Erfassung der Gebrauchstauglichkeit.....	157
8.2.4	Zusammenfassende Betrachtung des Methoden-Inventars.....	160
9	Analyse und Bewertung des Multi-Trait/Multi-Method-Ansatzes als Validierungsstrategie	164
10	Offene Forschungsfragen	169
10.1	Methodische Anpassungen	169
10.2	Anpassungen der Validierungsstrategie.....	171
11	Zusammenfassung.....	173
	Literaturverzeichnis.....	176
	Abkürzungsverzeichnis	187
	Abbildungsverzeichnis.....	188
	Tabellenverzeichnis	194
A	Orientierungshilfe zur Interpretation der Multi-Trait/Multi-Method-Matrizen.....	199
B	Multi-Trait/Multi-Method-Matrix der Machbarkeitsstudie	201
C	Multi-Trait/Multi-Method-Matrix der Re-Design-Studie	205
D	Multi-Trait/Multi-Method-Matrix der 3DA-Untersuchung	210
E	Multi-Trait/Multi-Method-Matrix der Sensor-Studie	215

ABSTRAKT

Bei der Neuentwicklung oder Anpassung von Funktionen für Flugzeug-Cockpits ist sicherzustellen, dass diese Änderungen weder die Aufrechterhaltung des Situationsbewusstseins der Nutzer gefährdet noch deren Arbeitsbelastung in kritischem Ausmaß erhöht, um die weitere Gebrauchstauglichkeit der Nutzerschnittstelle zu gewährleisten. Daher ist vor Implementierung einer neuen oder geänderter technischer Funktion deren Auswirkung auf den Nutzer im Hinblick auf diese interdependenten Qualitätskriterien zu überprüfen. Hierzu ist es notwendig, die Qualitätskriterien möglichst umfassend, das heißt anhand eines multidimensionalen Human-Factors-Methoden-Inventars, zu erfassen. Dies sollte aus ökonomischen Gründen bereits in einem frühen Stadium der Entwicklung in hierfür geeigneten Simulationsumgebungen unter Einbindung der Nutzer und unter Abbildung eines repräsentativen Aufgaben- und Nutzungskontextes geschehen. Dies hat zur Konsequenz, dass das Methoden-Inventar systematisch und unter Erhalt des Nutzungskontextes in die Evaluierungen der technischen Cockpit-Funktionen eingebunden werden muss.

In der vorliegenden Arbeit wird daher untersucht, wie ein Human-Factors-Methoden-Inventar beschaffen sein muss, um in eine spezifische Simulationsumgebung wie der eines Flugzeug-Cockpits eingebunden werden zu können. Hierzu werden Human-Factors-Methoden für Situationsbewusstsein, Workload und Gebrauchstauglichkeit sukzessive kombiniert, angepasst und in die Simulationsumgebung integriert. Anhand von vier technischen Funktionen wird das Methoden-Inventar zur Evaluierung der Mensch-Maschine-Schnittstelle Cockpit verwendet. Die Eignung des Methoden-Inventars wird anhand eines angepassten Multi-Trait/ Multi-Method-Ansatzes als Validierungsstrategie überprüft. Es zeigt sich, dass die untersuchten Qualitätskriterien anhand der eingesetzten Methoden nur schwer hinsichtlich konvergenter und diskriminanter Validität abzugrenzen sind. Eine Überschneidung der Konstrukte wird unter Einsatz der verwendeten Validierungsstrategie somit offenkundig.

Die Konsequenzen dieses Ergebnisses für die Weiterentwicklung des Instrumentariums zur Erfassung der Qualitätskriterien sowie der darauf aufbauenden Validierungsstrategien wird abschließend diskutiert.

ABSTRACT

When implementing technical functions into aircraft cockpits, it has to be ensured that these changes will not endanger the formation of operators' Situational Awareness, nor that they will increase operator workload beyond measure in order to warrant the cockpit's Usability. This is why a new or changed technical function has to be assessed regarding its impact on the user and according to the above mentioned interdependent quality criteria. Thus, a multidimensional inventory of Human Factors methods has to be applied. This should be done at an early stage of the function's development for economical reasons. For reasons of interpretability expert users in interaction with their tasks and in realistic contexts have to be studied. For an evaluation of technical functions the method inventory has to be integrated thoroughly into the simulation environment in order to preserve the user context.

In this thesis it is investigated, how a Human Factors method inventory should be compiled, for its implementation into a specific simulation environment for aircraft cockpits. Therefore Human Factors methods will be successively combined, adapted and integrated within the simulation environment. For evaluating the man machine interface, cockpit, four technical functions are used. The suitability of the method inventory is assessed on the basis of a validation strategy which is an adapted Multi-Trait/ Multi-Method approach. It is shown, that the evaluated quality criteria apparently overlap heavily and therefore cannot be well differentiated according to their convergent and discriminant validity. An overlap of these constructs is therefore apparent, which could be shown with the used validation strategy.

In conclusion the consequences of these results are discussed regarding the further developments of the quality criteria measures and its impact on further validation strategy.

1 EINLEITUNG

In einer sich verändernden Welt mit immer weiterführenden technischen Entwicklungen stellt sich die Frage, inwiefern der Mensch diese technischen Erweiterungen zielführend und für sich zufriedenstellend nutzen kann. Um diese Frage beantworten zu können, ist eine strukturierte Untersuchung von Mensch-Maschine-Schnittstellen (MMS), also der Bedienoberfläche, über die ein Mensch mit einem technischen System interagiert, und zwar im Hinblick auf die jeweilige Anwendungssituation und nicht nur allgemein notwendig. Dies trifft auf eine sehr breite Problemlage zu, die von der alltäglichen Bedienung von beispielsweise Haushaltsgeräten und Fahrzeugen bis hin zu sicherheitsrelevanten Mensch-Maschine-Interaktionen reicht, wie sie beispielsweise in der Nukleartechnik oder auch in der Luftfahrt vorliegen. Gerade in der Luftfahrt und somit auch in der Weiterentwicklung hochagiler Flugzeuge ist es notwendig, den sich verändernden Anforderungen gerecht zu werden, um die Sicherheit im Flug, aber auch am Boden zu gewährleisten. Zur Anpassung an die veränderten Gegebenheiten sind daher immer wieder Anpassungen bereits gegebener Flugsysteme anhand von Funktionserweiterungen notwendig.

Speziell sind Funktionserweiterungen an der MMS Cockpit bedeutsam, da darüber der Pilot mit dem Flugzeug interagiert (s.a. Kellerer, 2010). Daraus folgt, dass Funktionserweiterungen des Cockpits so implementiert werden müssen, dass das bisher gegebene Zusammenspiel von Mensch und Maschine weiterhin reibungslos funktioniert und dem Menschen mit den neuen Funktionalitäten eine zeitnahe und zielführende Bearbeitung seiner Aufgaben ermöglicht wird. Das heißt, dass der Pilot, der als Entscheidungsträger in einem komplexen System agiert, nicht über die Maße belastet werden soll, damit er mit den ihm verfügbaren, aufgabenspezifischen Ressourcen in den sich bietenden Situationen adäquat reagieren kann. Daher müssen Funktionserweiterungen vor ihrer Implementierung in ein Cockpit hinsichtlich ihres Einflusses auf die Arbeitsbelastung und bezüglich der Aufrechterhaltung des Situationsbewusstseins des Piloten geprüft werden, um zudem eine gebrauchstaugliche Bedienung sicherzustellen (Billeri, Kellerer & Zimmer, 2018).

Bei Airbus Defence and Space ist das Team für Human Factors Engineering and Flight Deck (GE) unter der Leitung von Dr. Peter Sandl dafür zuständig, die ergonomische Bedienung von Cockpit-Funktionen zu entwickeln und in bestehende Systeme unter den Aspekten der Gebrauchstauglichkeit zu integrieren. In diesem Arbeitsumfeld wurde ein Methoden-Inventar, bestehend aus qualitativen und quantitativen Methoden entwickelt, welche Cockpit-Funktionen in einer Simulationsumgebung im Hinblick auf deren Auswirkungen auf die Bildung und Aufrechterhaltung von Situationsbewusstsein, Nutzer-Workload sowie Gebrauchstauglichkeit bewertet.

1.1 Problemstellung

Für die Evaluationsumgebung eines Flugzeug-Cockpit-Simulators wird ein valides, methodisches Rahmenkonzept benötigt. Mit Hilfe dieses Rahmenkonzeptes soll es durch die gezielte Anwendung von subjektiven und objektiven Methoden gelingen, Änderungen im Cockpit hinsichtlich ihres Ein-

1 Einleitung

flusses auf das Situationsbewusstsein von Experten sowie deren empfundener Workload zu untersuchen, um so die Gebrauchstauglichkeit der Cockpit-Modifikation zu bestimmen. Um sicherzustellen, dass das methodische Rahmenkonzept die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit erfasst und für die Daten-Erhebung in der genannten Versuchsumgebung geeignet ist, soll dieses auch abschließend validiert und in der Folge angepasst werden.

1.2 Ziele

Aus der beschriebenen Problemstellung lassen sich vier Teilziele ableiten, die erreicht werden müssen, um ein multidimensionales Methoden-Inventar als Rahmenkonzept in die Cockpit-Simulationsumgebung einzubinden und dessen Eignung festzustellen.

1. Erstellung eines Methoden-Inventars zur Erhebung der Qualitätskriterien und dessen Integration in die Cockpit-Simulationsumgebung
2. Evaluierung von technischen Funktionen in der Cockpit-Simulationsumgebung unter Anwendung des Methoden-Inventars
3. Eignungsbewertung des Methoden-Inventars durch Verwendung einer geeigneten Validierungsstrategie
4. Ableitung von Handlungsempfehlungen

Ein schematisches Vorgehen zur Erreichung der vier Teilziele wird in Abbildung 1.1 dargestellt.

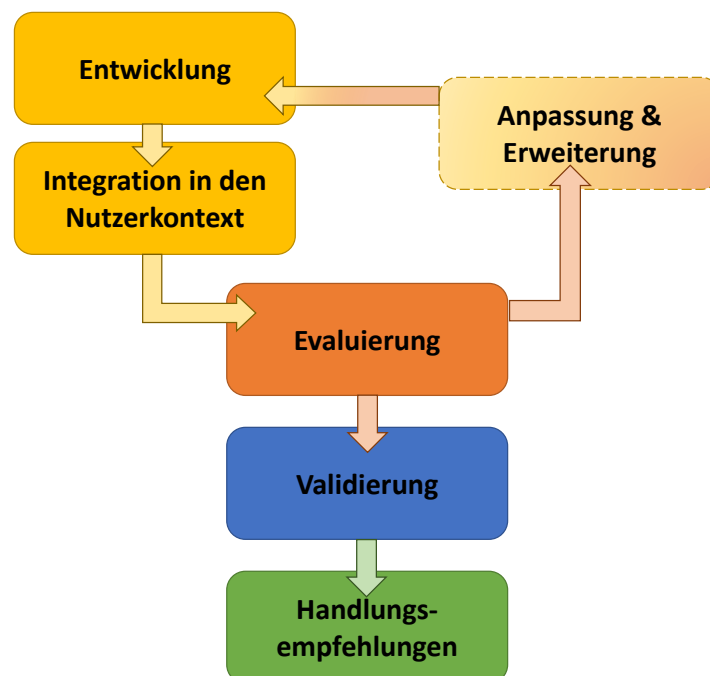


Abbildung 1.1 Vorgehen zur Implementierung und Eignungsfeststellung eines Methoden-Inventars

Unter Betrachtung des oben aufgeführten Ablaufdiagramms werden die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit in einem ersten Schritt als theoretische Kon-

1 Einleitung

zepte analysiert. Vorhandene Methoden aus dem Bereich Human-Factors werden auf dieser Grundlage ausgewählt, zu einem Inventar zusammengefügt und in den jeweiligen Nutzungskontext der technischen Funktionsveränderungen im Cockpit eingebunden (Teilziel 1). Mit Hilfe der technischen Cockpit-Modifikationen wird dann das vorhandene Methoden-Inventar in der Simulationsumgebung evaluiert. Zu diesem Zweck werden verschiedene Funktionserneuerungen in unterschiedlichen Entwicklungsstufen in den Cockpit-Simulator implementiert und in repräsentativen Situationen mit den hierfür geeigneten subjektiven und objektiven Methoden des Methoden-Inventars evaluiert (Morgenstern, Kellerer, Neujahr & Zimmer, 2016). Im Hinblick auf die einzeln durchgeführten Evaluierungen wird das Methoden-Inventar iterativ ergänzt und angepasst und erneut in folgende Evaluierungen überführt (Teilziel 2). Die integrative Validierung dieser Methoden erfolgt unter Zuhilfenahme eines Multi-Trait/ Multi-Method (MTMM)-Ansatzes nach Campbell & Fiske, 1959, als Validierungsstrategie (Teilziel 3). Dieses Vorgehen erlaubt es, in einem abschließenden Schritt Handlungsempfehlungen (Teilziel 4) hinsichtlich der weiteren Anwendung der verwendeten Methoden, deren Eignung und möglicher Anpassungen abzuleiten.

1.3 Aufbau der Arbeit

Der Aufbau der Arbeit orientiert sich an den in Kapitel 1.2 erwähnten vier Teilzielen zur Entwicklung, Evaluierung und Validierung eines Methoden-Inventars in einer Cockpit-Simulationsumgebung unter Ableitung von weiteren Handlungsanweisungen. In Kapitel 2 wird der praktische Hintergrund erläutert, warum es notwendig ist, Cockpit-Modifikationen im Nutzerkontext mit geeigneten Bewertungskriterien möglichst früh im Entwicklungsprozess – wie beispielsweise in einer Simulationsumgebung – zu bewerten. Die in diesem Kontext in Kapitel 2 abgeleiteten Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit und deren theoretischer Hintergrund werden in Kapitel 3 erarbeitet und gängige Messmethoden dieser Qualitätskriterien dargestellt. Am Ende dieses Kapitels wird somit das in der Simulationsumgebung verwendete Methoden-Inventar vorgestellt. Wie dieses Methoden-Inventar schließlich in die vorhandene Simulationsumgebung eingebunden werden kann, wird in Kapitel 4 erörtert. In Kapitel 5 werden die praktisch durchgeführten Evaluierungen des Methoden-Inventars erläutert und die Evaluierung an einer konkreten Cockpit-Modifikation exemplarisch aufgearbeitet. Der theoretische Hintergrund zur Beschreibung der eingesetzten Validierungsstrategie im Rahmen des Multi-Trait/ Multi-Method-Ansatzes wird in Kapitel 6 beschrieben. Folglich werden die konkreten Ergebnisse der Validierung des eingesetzten Methoden-Inventars in Kapitel 7 offengelegt und schließlich in Kapitel 8 diskutiert. Ergänzend werden konkrete Handlungsempfehlungen für die Weiterverwendung der eingesetzten Methoden des Methoden-Inventars abgeleitet und erörtert. Auch die Eignung der Validierungsstrategie wird abschließend in Kapitel 9 diskutiert. Mögliche Anpassungen und Vorschläge zu weiteren Untersuchungen des Methoden-Inventars und der Validierungsstrategie werden in Kapitel 10 vorgeschlagen. Zum Abschluss wird in Kapitel 11 das zu Grunde liegende Vorgehen zur Evaluierung und Validierung des Methoden-Inventars sowie die maßgeblichen Erkenntnisse zusammengefasst.

1 Einleitung

Bezugnehmend auf die in Kapitel 1.2 formulierten Teilziele bedeutet dies, dass in Kapitel 2 bis 4 die Entwicklung und Integration des Methoden-Inventars beschrieben wird, während in Kapitel 5 die Evaluierung des Methoden-Inventars anhand von vier technischen Cockpit-Modifikationen verdeutlicht und anhand eines Beispiels die Datengrundlage der Evaluierungen erklärt wird. In Kapitel 6 und 7 werden das Vorgehen der Validierungsstrategie sowie die hieraus abzuleitenden Ergebnisse dargestellt. Die zu folgernden Handlungsempfehlungen für das Methoden-Inventar sowie hinsichtlich der Validierungsstrategie können auf Grund der in Kapitel 8 und 9 geführten Diskussion abgeleitet werden und werden in Kapitel 10 in mögliche weiterführende Forschungsfragen überführt.

2 ERHEBUNG UND BEWERTUNG DER MMS COCKPIT IN EINEM KOMPLEXEN SOZIO-TECHNISCHEN SYSTEM

Human-Factors-Analysen stellen eine Disziplin dar, in welcher die Verknüpfung von technischen sowie menschlichen Faktoren im Fokus steht (Badke-Schaub, Hofinger & Lauche, 2008); eine Disziplin also, die sich mit dem Design und der Evaluierung von Aufgaben, Arbeitsumgebungen und Systemen befasst und zudem die Fähigkeiten, Fertigkeiten und Grenzen des Menschen betrachtet (Karwowski, 2012). In dieser Disziplin sind sozio-technische Systeme betroffen, in denen Menschen, Sozialstrukturen und technische Elemente miteinander interagieren und diese Interaktion so gestaltet werden muss, dass die Ziele eines Systems erreicht werden können (Czaja & Sankaran, 2012).

Die Bedienung von Mensch-Maschine-Schnittstellen (MMS) ist daher zentraler Teil eines sozio-technischen Systems, welches in einem spezifischen situativen Nutzerkontext stattfindet. Folglich ist die Auslegung, Entwicklung und Anpassung von MMS an die Anforderungen und Bedürfnisse der Nutzer und auch die Qualitätsbestimmung der MMS in einem sozio-technischen System auszurichten. Die MMS selbst ist hierbei ein Mittel zur Erreichung des Systemzieles in dem jeweiligen Nutzungskontext, in welchem sie bedient wird.

Die Beschreibung des Nutzungskontextes einer MMS ist daher grundlegend, um die MMS weiterzuentwickeln und schließlich die Qualität einer MMS bewerten zu können. Gerade bei sicherheitsrelevanten MMS wie einem Flugzeug-Cockpit, das in ein komplexes sozio-technisches System eingebettet ist, müssen deshalb die Kriterien, mit denen die Qualität der Schnittstelle bewertet werden sollen sowie die Methoden, welche die Kriterien erfassen, sorgfältig ausgewählt werden, damit sie zu den situativen Anforderungen passen.

2.1 Anpassungen der MMS Flugzeug-Cockpit

Der funktionale Leistungsumfang von hochagilen Luftfahrzeugen wird während ihrer langjährigen Nutzung fortwährend erweitert. Diese Änderungen werden durch das Cockpit für den Nutzer, d.h. dem Piloten, nutzbar und müssen bei deren Anwendung einen zugleich effizienten und sicheren Flugbetrieb ermöglichen (Kellerer, 2010). Diese Erweiterungen erfolgen oft über eine Anpassung des Systems und dessen Bedienung, beispielsweise durch die Erweiterung von Software, die der Bedienung hinterlegt ist. Die Notwendigkeit, das technische System und somit das Cockpit zu modifizieren, entsteht im vorliegenden Fall zumeist aus einem Bedürfnis von Betreibern oder auch Nutzern heraus, eine definierte Erweiterung oder Verbesserung des Fähigkeitsspektrums zu erzielen, um bestimmte Nutzerziele im komplexen sozio-technischen System zu erreichen. Dieses Bedürfnis wird im folgenden Entwicklungsprozess in Anforderungen umformuliert, die die zu entwickelnde technische Lösung erfüllen muss, um die Bedürfnisse des Nutzers zufrieden zu stellen. Diese Anforderungsanalyse ist oft in ein Vorgehensmodell wie dem bekannten V-Modell XT (beispielsweise Version 2.3 des Bundesministeriums des Innern, für Bau und Heimat) eingebunden, welches zur Planung und Durchführung von IT-System-Entwicklungsprojekten verwendet wird. Ein Vorgehen nach dem V-Modell wird als

Grundlage zur Kommunikation für alle Stakeholder genutzt, um Projektrisiken zu minimieren, die Gesamtkosten zu verringern und um die Qualität des Projektes sicherzustellen. Das V-Modell XT ist durch seine modulare Aufbauweise an verschiedene System-Entwicklungsprojekte flexibel anpassbar und legt die Definition der Anforderungen als ersten Entscheidungspunkt zur weiteren Systemerstellung fest (Angermeier et al., 2006). Unter Anforderungen versteht man hierbei die Fähigkeiten, die beispielsweise die zu entwickelnde Erweiterung / technische Funktion erfüllen muss, damit der Nutzer zukünftig das von ihm angestrebte Ziel erreichen kann. Diese Anforderungen werden formell vorgegeben und nachvollziehbar dokumentiert (Institute of Electrical and Electronics Engineers, 1990). Die Anforderungen werden schließlich in einer Anforderungsspezifikation gebündelt, die wiederum vollständig, konsistent, erschwinglich (d.h. umsetzbar) und abgegrenzt sein soll (Rupp, 2014). Die Festlegung der Anforderungen ist für die Entwicklung der Funktionsveränderungen im Cockpit wesentlich, da so der Lösungsrahmen der zu entwickelnden technischen Funktion abgesteckt wird. Um das Risiko von Fehlentwicklungen zu reduzieren und um die Performanz des ersten Designs der technischen Funktion und dessen Vollständigkeit zu testen, wird die Modifikation der MMS Cockpit bereits im Entwicklungsprozess, d.h. vor deren Integration in ein Teil- oder dem Gesamt-System, überprüft. Mit dieser Vorgehensweise soll sichergestellt werden, dass das bisherige Design einen bestimmten Grad an Ausgereiftheit nachweist und festgestellt werden, ob und an welchen Aspekten weitere Verbesserungen vorgenommen werden müssen oder ob vielleicht sogar ein völliger Neuentwurf des bisherigen Designs notwendig ist. Hierzu stehen je nach Kosten-Nutzen-Abwägung mehrere Überprüfungsvorgehen zur Verfügung, in deren Fokus aber jeweils die Befragung der Nutzer und dessen Nutzungskontext steht.

Das Entwicklungsvorgehen hat somit zur Folge, dass auch die formale Beschreibung des Systems und der hierin zu implementierenden technischen Funktionen nur mit Fokus auf den Nutzer und der jeweiligen Handlungssituation - dem Nutzerkontext – geschehen kann.

2.2 Der Nutzer und dessen Nutzungskontext

Bei der Gestaltung von Mensch-Maschine-Schnittstellen in einem komplexen sozio-technischen System wird unter Betrachtung der menschlichen Faktoren darauf geachtet, dass das Zusammenspiel von Mensch und Maschine optimiert wird (Czaja & Sankaran, 2012). Dies kann aber nur geschehen, wenn die MMS auch an die Anforderungen und Bedürfnissen des Nutzers, dessen Aufgaben, seiner Arbeitsumgebung sowie an die systembezogene Umwelt angepasst ist. Eine maßgebliche Anforderung an die vom Nutzer benötigten Cockpit-Funktionen ist es daher, dass der Pilot die Informationen, die er zu einem Zeitpunkt braucht, in einer geeigneten Form zur Verfügung gestellt bekommt. Informationen müssen daher schnell verfügbar, leicht verständlich, gut sichtbar und in passender Menge dargestellt werden. Die geforderten Cockpit-Funktionen können hierbei sowohl in Form von veränderten Informationsdarstellungen auf unterschiedlichen Displays auftreten als auch verschiedene Kontrollfunktionen betreffen, die eine veränderte Bedienung zur Folge haben. In diesem Zusammenhang impliziert das Funktionsdesign die planerische Festlegung wie, wann und wo Informationen dargestellt werden und in welcher Art und Weise der Nutzer mit dem System interagieren soll. Ein

2 Erhebung und Bewertung der MMS Cockpit in einem komplexen sozio-technischen System

erster Entwurf des Funktionsdesigns kann dann durch den Nutzer beispielsweise anhand von Stift und Papier, anhand von Präsentationen am Computer, durch ein Vorführmodell oder Mock-Up als auch durch einen Simulator überprüft und bewertet werden. Welches Mittel zur Überprüfung gewählt wird, ist sowohl von einer Kosten-Nutzen-Analyse als auch von der Komplexität der Arbeitsumgebung abhängig. Da es sich bei den zu testenden Funktionalitäten im Cockpit meist um komplexe und umfangreiche Änderungen handelt, ist eine Überprüfung in einem realistischen Arbeitsumfeld, d.h. in einer Simulationsumgebung, angezeigt.

Das heißt im Rückschluss auch, dass eine Bewertung der Cockpit-Modifikation in der realistischen Arbeitsumgebung nur anhand von Nutzern vorgenommen werden kann, die über die entsprechende Expertise verfügen. Experten – als Personen mit Expertise – zeichnen sich nach Krems (1997) dadurch aus, dass sie die ihnen zugewiesenen Aufgaben effizienter, d.h. mit einem geringeren Ressourceneinsatz wie Zeit und in höherer Qualität, d.h. bei einer geringeren Fehlerquote, bearbeiten können als Nicht-Experten. Das Expertentum selbst entsteht hierbei durch fundierte bereichsspezifische Kenntnisse, die auf mehreren Jahren an Erfahrung beruhen. Dieses Expertenwissen ist für Piloten hochagiler Luftfahrzeuge deshalb wesentlich, da diese schnell situationsangemessene Entscheidungen treffen müssen. Für die empirische Analyse der MMS Cockpit hat dies zur Konsequenz, dass die Expertise von Versuchspersonen von entscheidender Bedeutung ist.

Mit einem zunehmenden Leistungsumfang des Flugzeuges ist es allerdings notwendig, den Piloten bei seiner Tätigkeit als ‚Entscheider‘ durch beispielsweise Automatisierung so zu unterstützen, dass eine effektive Auslegung der Mensch-Maschine-Kombination möglich ist. Dies wiederum erfordert ein kohärentes Cockpitdesign (Wilkinson, 1992). Zur Überprüfung, ob diese Kohärenz fortbesteht, wenn Änderungen in das Cockpit integriert werden, muss die Komplexität des Handlungskontextes des Piloten innerhalb der Simulationsumgebung abgebildet (Morgenstern et al., 2016) und müssen die Cockpit-Modifikationen von Domänenexperten hierin bewertet werden. Das Aufgabenspektrum eines Piloten für hochagile Luftfahrzeuge umfasst neben den vier übergeordneten Flugaufgaben wie Flugwegkontrolle, Navigation, Kommunikation und Systemmanagement (Eichinger, 2011; Wickens, 2003) auch Taktik/ Bedrohungsmanagement (Eichinger, 2011; Kellerer, 2010). Alle übergeordneten Flugaufgaben nach Wickens (2003), jedoch insbesondere die Aufgaben des Taktik/ Bedrohungsmanagements, erfordern hier die Expertise von Domänenexperten. Domänenexperten in diesem Zusammenhang sind Piloten, die das zu untersuchende Flugzeugmuster professionell nutzen und außerdem auch die Ausbildung und Erfahrung bezüglich der spezifischen Aufgabenstellungen des Taktik- und Bedrohungsmanagements besitzen. Da diese Domänenexperten die Cockpit-Modifikationen eines Flugzeugmusters in einer somit spezifischen Simulationsumgebung bewerten sollen, ist die Nutzergruppe einer ohnehin schwer zu erreichenden, kleinen Zielgruppe (s.a. Rabl, 2016) eingeschränkt. Dies wiederum erfordert ein extensives Untersuchen der zur Verfügung stehenden Stichprobe aus ökonomischen Gründen und aus Gründen der Effizienz – beispielsweise in Form eines umfangreichen Within-Subject-Experimentaldesigns. Dies ist weiterhin deshalb zulässig, da es sich um Designbewertungen von Cockpit-Modifikationen handelt. So stellen Nielsen und Landauer (1993) fest, dass die Entdeckung von Problemen der Gebrauchstauglichkeit anhand der Durchführung von

Nutzerstudien zur Untersuchung von Nutzerschnittstellen Poisson-verteilt ist. Allerdings ist auch bei kleinen Nutzergruppen insbesondere auf deren Repräsentativität zu achten (Sauro & Lewis, 2012). Zusammenfassen lässt sich dies dahingehend, dass die meisten Probleme der Gebrauchstauglichkeit bereits von drei bis fünf befragten Nutzern erkannt werden und weitere Befragungen nur zu wenigen zusätzlichen Erkenntnissen führen. Den größten Nutzen hingegen erreicht man durch ein iteratives Vorgehen anhand kleiner Nutzerstudien (C. W. Turner, Lewis & Nielsen, 2006). Da die Durchführung der Studien mit Domänenexperten schon in einem frühen Stadium der Designentwicklung (wie in Kapitel 2.1 beschrieben) zur Risikominimierung durchgeführt werden, ist ein iteratives Vorgehen immanent. Die im Rahmen der Nutzerstudie erkannten und in der Interaktion mit der Nutzerschnittstelle entdeckten Probleme werden daher aufgearbeitet und in ein weiteres Design überführt, bevor dieses erneut bewertet werden kann. Um die Iterationsanzahl jedoch aus Gründen der Effizienz so gering wie möglich zu halten, ist eine eingehende Überprüfung der Cockpit-Modifikation anhand der Domänenexperten in deren Nutzungsumgebung, wie sie beispielsweise durch einen Cockpit-Simulator abgebildet wird, notwendig.

2.3 Die Überprüfung von Cockpit-Modifikationen in einem Cockpit-Simulator

Zur Überprüfung veränderter Cockpit-Funktionen in einer realistischen Arbeitsumgebung stehen bei Airbus Defence and Space drei Räumlichkeiten zur Verfügung: 1) der (De-) Briefing-Raum, 2) der Kontrollraum und 3) der Cockpit-Simulator.

Im (De-) Briefing-Raum (1) wird dem Piloten mit den üblichen Präsentationsmöglichkeiten die zu untersuchende Funktionalität vorgestellt und der Ablauf des Versuchs sowie das Geschehen innerhalb der Simulation beschrieben. Des Weiteren bietet dieser Raum die Möglichkeit eines De-Briefings. Ein De-Briefing bietet hier die Möglichkeit, die Ereignisse des Versuchs nach dessen Abschluss nochmals zusammenzufassen und aufzubereiten. Videomaterial, das während der Untersuchung aufgezeichnet wurde, kann dazu genutzt werden, eine Videokonfrontation durchzuführen, um so den Piloten hinweisgestützt zu bestimmten Ereignissen während der Simulation zu interviewen.

Ein weiterer Raum der Versuchsumgebung ist der Kontrollraum (2). Er dient der Ablaufüberwachung der Evaluierung und deren Kontrolle, der Beobachtung der Ereignisse im Cockpit und ist ein ergänzendes Mittel der Datenaufzeichnung. In diesem Raum befinden sich während des Versuchs jene Personen, die das Versuchsszenario mitgestalten, steuern sowie die Handlungen im Cockpit beobachten. Den Versuchsleitern, Konföderierten oder weiteren Beteiligten ist es möglich, während des Versuchs die Tastatur-Eingaben des Piloten sowie dessen Blickvektor in Echtzeit zu beobachten und relevante Ereignisse zeitlich im Simulations-Geschehen zu markieren und zu kommentieren, die später anhand der erwähnten Videokonfrontation im De-Briefing analysiert werden können.

Das Herzstück der Messumgebung ist der Cockpit-Simulator (3), wie in Abbildung 2.1 dargestellt. Der Cockpit-Simulator ist ein Full-Replica Cockpit, welches dem realen Flugzeug hinsichtlich Geometrie, Anzeige, Bedienelementen und Haptik nahekkommt. Das Gefühl, sich in einer realistischen Handlungssituation zu befinden, wird dem Piloten durch eine sphärische Außenansicht mit $\pm 135^\circ$ Azimut

2 Erhebung und Bewertung der MMS Cockpit in einem komplexen sozio-technischen System

sowie +90°/ -45° vertikaler Sicht in einem 2 m Radius vermittelt. Ein spezieller Vorteil des Simulators ist dessen Zugänglichkeit für ‚Rapid-Prototyping‘. Das heißt, dass neue Cockpit-Funktionen hier schnell implementiert und überprüft werden können, da keine originale Hardware und Software verwendet wird.



Abbildung 2.1 Darstellung des Cockpit-Simulators

Zusammengenommen ergeben alle drei Räumlichkeiten die Simulationsumgebung, die dazu genutzt wird, um die Cockpit-Modifikationen anhand von Domänenexperten zu evaluieren. Neben den Experten und einer für sie zur Bewertung der Cockpit-Modifikation geeigneten, realistischen und repräsentativen Arbeitsumgebung, sind aber weiterhin geeignete Bewertungskriterien und Methoden notwendig, um die Qualität der geänderten MMS zu überprüfen.

2.4 Bewertungskriterien und deren Erhebung im Cockpit-Simulator

Zum Nachweis der Eignung der Cockpit-Modifikation ist es notwendig, dass die aus den Bedürfnissen der Nutzer entstandenen Anforderungen so umgesetzt worden sind, dass der Nutzer seine Ziele in dem komplexen sozio-technischen System, in dem er agiert, erfolgreich erreichen kann. Das heißt, um zu definieren, wann eine Cockpit-Modifikation geeignet ist, müssen die Nutzer, die Nutzerziele und der Nutzerkontext, in dem diese Ziele erreicht werden sollen, bekannt sein (ISO 9241-11:2018(E)). Somit ist im Rückschluss die Eignung der Cockpit-Modifikation nur dann festzustellen, wenn die (technische) Auslegung der Cockpit-Modifikation bekannt ist und welches Ziel der Pilot eines hochagilen Flugzeuges in einer bestimmten Nutzungsumgebung oder -situation erreichen möchte. Die Aufgabe der Nutzerschnittstelle ist es in diesem Zusammenhang, den Dialog zwischen Mensch und Maschine zu gestalten, damit der Nutzer sein Handlungsziel erreicht. Der Dialog beinhaltet somit, dass der Nutzer zum einen mittels Eingabegeräte das System steuert und zum anderen vom System vorverarbeitete Informationen bekommt (Clamann & Kaber, 2004). Die Nutzerschnittstelle eines hochagilen Flugzeuges stellt daher die Informationen in einer bestimmten Geschwindigkeit und Qualität zur Verfügung, die der Pilot benötigt, um die Prozesse durch Bedienung auszuführen (Calhoun, 2000), die seinem Nutzerziel dienen. Diese Interaktion liefert den entscheidenden Beitrag zur globalen Effektivität eines komplexen Flugsystems, da diese für operationelle Agilität sorgt – also den Piloten die Fähigkeit vermittelt, adäquat zu reagieren und so den Missionserfolg sicherzustellen (Le Blaye, 2000). Dies ist allerdings für den Piloten in seiner Arbeitswelt und auf Grund der

2 Erhebung und Bewertung der MMS Cockpit in einem komplexen sozio-technischen System

Leistungserweiterungen von Subsystemen zunehmend erschwert und erlaubt eine Bedienung des Gesamt-Systems in komplexen Situationen nur, wenn dessen Arbeitsbelastung durch eine entsprechende Cockpit-Auslegung handhabbar gestaltet wird, damit dieser in seiner Rolle als Entscheidungsträger agieren kann (Wilkinson, 1992). Die MMS ist also dann für den Piloten nutzbar, wenn die Informationen, die der Pilot zum Treffen seiner Entscheidungen benötigt, zur richtigen Zeit, in geeigneter Form und in der richtigen Menge zur Verfügung gestellt werden, ohne dessen Arbeitsbelastung in kritischem Maße zu erhöhen. Zusammengefasst heißt das, dass die Cockpit-Modifikation dann gebrauchstauglich ist, wenn diese das Situationsbewusstsein des Nutzers stützt, ohne die Workload in einem Maße zu erhöhen, dass ein erfolgreiches Agieren des Experten in dem komplexen sozio-technischen System nicht mehr möglich ist (Morgenstern et al., 2016). Inwiefern diese Gebrauchstauglichkeit der Cockpit-Modifikation besteht, muss sorgfältig und möglichst früh im Entwicklungsprozess überprüft werden, um das Risiko einer Fehlentwicklung zu minimieren. Daraus ist zu folgern, dass Situationsbewusstsein, Workload und Gebrauchstauglichkeit für einen erfolgreichen Mensch-Maschine-Dialog essentiell sind und deshalb als Bewertungskriterien für die Cockpit-Modifikation herangezogen werden. Clamann und Kaber (2004) identifizieren folgende Voraussetzung für die Erfassung von Gebrauchstauglichkeit als übergeordnetes Konstrukt und die hierzu verwendeten Methoden in der Luftfahrt:

- Schnelle Durchführbarkeit
- Kosteneffektivität
- Einbindung in den Lebenszyklus des Produktes
- Input von Domänenexperten
- Übertragbarkeit und Skalierbarkeit auf das Cockpit

Hinter diesen Voraussetzungen steht in erster Linie der Gedanke, dass die verwendeten Methoden als Hilfsmittel im Entwicklungsprozess von Cockpit-Modifikationen gedacht sind und daher den Gegebenheiten anzupassen sind. Demnach soll durch die verwendeten Methoden weder die Weiterentwicklung des Systems verlangsamt oder die Entwicklungskosten erhöht werden. Die Methoden sollen zudem nicht nur auf einen einzelnen zu evaluierenden Aspekt anwendbar sein, sondern auch für zusammenhängende Komponenten eines Systems. Außerdem ist es von Bedeutung, Gebrauchstauglichkeit nicht nur einmal und punktuell zu bestimmen, sondern deren Bewertung in den gesamten Entwicklungsprozess zu integrieren. Zentral ist jedoch die Einbindung der Domänenexperten und deren Wissen und Erfahrung als essentieller Beitrag für den Entwicklungsprozess (Coombs, 1999). Im Hinblick auf diese nutzerzentrierte Herangehensweise ist es notwendig, ein Methoden-Inventar zu verwenden, das dieses Wissen in geeigneter Form erfassen kann, um hiervon auch in der Weiterentwicklung der Cockpit-Modifikation zu profitieren. Eine Ausrichtung der Methoden an die Experten-Stichprobe ist somit erforderlich (Morgenstern et al., 2016).

2 Erhebung und Bewertung der MMS Cockpit in einem komplexen sozio-technischen System

Zusammenfassend lässt sich also bemerken, dass zur Überprüfung der Qualität der Nutzerschnittstelle die Bewertungskriterien Gebrauchstauglichkeit, Situationsbewusstsein und Workload im Cockpit-Simulator zuverlässig erfasst werden müssen. Hierzu ist ein Konvolut aus qualitativen und quantitativen Methoden notwendig, das an der Experten-Stichprobe ausgerichtet ist, sich für eine Anwendung im Cockpit-Simulator eignet – also die Repräsentativität des Handlungskontextes erhalten und weiterhin ökonomisch sinnvoll eingesetzt werden kann.

3 KONSTRUKTE UND METHODEN ZUR QUALITÄTSBESTIMMUNG DER MMS FLUGZEUG-COCKPIT

Sind die Anforderungen an die technische Änderung definiert und umgesetzt, müssen die Auswirkungen der Änderungen überprüft werden, da diese sich direkt auf die Qualität der Mensch-Maschine-Interaktion (MMI) im Cockpit auswirken. So sollte trotz der Änderungen im Cockpit sichergestellt werden, dass dem Piloten alle relevanten Informationen zur Verfügung gestellt werden, dass dessen Arbeitsbelastung nicht über die Gebühr beansprucht wird und dass die Mensch-Maschine-Schnittstelle (MMS) zudem zielführend genutzt werden kann. Daher erfolgt die Überprüfung der MMI anhand der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit. Der theoretische Hintergrund dieser Qualitätskriterien wird deshalb in diesem Kapitel nicht umfänglich, sondern im Hinblick auf die einzuführenden Methoden des multidimensionalen Methoden-Inventars erarbeitet.

3.1 Situationsbewusstsein

Das Situationsbewusstsein als Beschreibung für ein adaptives, nach Außen gerichtetes Bewusstsein (Smith & Hancock, 1994) ist für die Luftfahrt im Allgemeinen von großer Bedeutung, da die augenblickliche Ausprägung des Situationsbewusstseins die Leistung des Piloten beeinflusst (Endsley, 1999). Der Begriff ist in der Luftfahrt schon lange vertreten (s.a. Coombs, 1999) und ist des Weiteren auch als Konzept hier auch ursprünglich zu verankern (s.a. Rabl, 2016). Bereits nach Ende des ersten Weltkrieges wurde Situationsbewusstsein als Voraussetzung für das erfolgreiche Durchführen einer Mission angesehen (Zimmer, 2016). Historisch gesehen werden in der militärischen Luftfahrt laut Coombs (1999) drei Aspekte deutlich, die das Situationsbewusstsein des Piloten beeinträchtigen: der Mangel an Informationen, ein Überfluss an Informationen sowie inadäquat dargestellte Informationen. Folglich ist Situationsbewusstsein als Designziel ein Kriterium, dass sowohl zur Bewertung neuer Technologien oder Designs genutzt wie auch zur Analyse von bestehenden operationellen Systemen (Stanton, Salmon, Walker, Baber & Daniel, 2011) als auch zur Überprüfung der sicheren Handhabung von Flugzeugen im Konsens zu den jeweiligen Zielen des Operateurs (Endsley, 1995).

Mica Endsley (1988a) formuliert den Begriff Situationsbewusstsein in den 80er Jahren zu einem Modell aus und prägt seit dem durch folgende Definition:

“... the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future.”

In dieser Definition sind die drei Aspekte (1) Wahrnehmung der relevanten Umweltelemente, (2) Verständnis der augenblicklichen Situation und (3) Projektion der derzeitigen Situation in die Zukunft maßgeblich, da dieser Dreisprung das Entscheiden und Handeln eines Menschen in seiner Umwelt bestimmt. Weitere individuelle Einflussfaktoren, die das Situationsbewusstsein moderieren, sind Fähigkeiten, Erfahrungen, Training und bestehende Schemata seitens des Akteurs. Aufgaben- und Systemfaktoren wie die Komplexität der Aufgabe und folglich auch die augenblickliche Arbeitsbelas-

tung, der Automatisierungsgrad des Systems oder das Design der Bedienoberfläche haben ebenso Einfluss auf Entwicklung und Erhalt des Situationsbewusstseins (Endsley, 1988a). Das Situationsbewusstsein wird durch die vorliegende Definition von den zu treffenden Entscheidungen des Handelnden abgegrenzt, da diese erst auf Grundlage des Situationsbewusstseins und in Abhängigkeit von dessen Qualität getroffen werden können (Endsley, 2000b).

Bei der Betrachtung des Konstruktes Situationsbewusstsein erweisen sich der zeitliche Verlauf und die sich hierin entfaltende situative Komplexität von erheblicher Bedeutung. Situationsbewusstsein wird anhand dieser Ergänzung immer noch als eigenständiges Konstrukt betrachtet, der Fokus wird hier jedoch auf ein kontinuierliches Situations-Monitoring des Nutzers im Hinblick auf die Veränderlichkeit der situativen Ereignisse gelegt (Sarter & Woods, 1991). Es wird kritisiert, dass die Definition von Situationsbewusstsein nach Endsley (1988a) eher statisch zu sehen ist, da Situationsbewusstsein als Ergebnis oder Produkt des oben erwähnten Dreisprungs zu sehen ist. Demnach wird eine sich ständig verändernde Umwelt und die hieraus resultierenden Anpassungsstrategien des Nutzers außer Acht gelassen – ein Disput, der sich auch in den vorgeschlagenen und entwickelten Methoden zur Erfassung des Situationsbewusstseins widerspiegelt (Durso & Sethumadhavan, 2008). Eine alternative Definition nach Smith und Hancock (1994), die Situationsbewusstsein eher als Prozess sehen, basiert auf dem Perceptual Cycle nach Neisser aus dem Jahr 1976. Neisser (1978) beschreibt Wahrnehmung als konstruktiven Prozess, in welchem ein Vorgehen – eine Strategie – zur weiteren Informationsgewinnung in der Umwelt generiert wird. Die gesammelten Informationen wiederum verändern die beim Nutzer vorhandenen Schemata, die dieser gezielt als Grundlage nutzt, um als Handelnder nach Informationen zu suchen, indem er seine sensorische Wahrnehmung danach ausrichtet. Smith und Hancock (1994) erweitern den Perceptual Cycle eben um die erwähnten Anpassungsstrategien des Nutzers und entwickeln so eine Verbindung zwischen der Information, die von der Umwelt zur Verfügung gestellt wird, der Wissensgrundlage, die der Nutzer zur Verwendung der Information benötigt und den Handlungen, mit denen er die Umwelt verändert, um seine Ziele zu erreichen. Auf Grundlage des von Smith und Hancock vorgeschlagenen Modells umfasst Situationsbewusstsein sowohl die kognitiven Prozesse, um dieses zu erwerben und zu erhalten sowie das sich ständig aktualisierende Produkt Situationsbewusstsein, das in der Interaktion mit der Umwelt entsteht (Stanton et al., 2011). Zimmer (2016) ergänzt den Perceptual Cycle nach Neisser anhand eines Doppel-Schleifen-Modells zur Durchführung dynamisch räumlicher Entscheidungen und Handlungen. Der Autor argumentiert, dass geplante Handlungen und Entscheidungen in einer räumlichen Situation bei einer Passung von Nutzererwartung und den zur Verfügung stehenden Situationsparametern an Genauigkeit und Geschwindigkeit gewinnen, während der Nutzer bei einer wahrgenommenen Diskrepanz von Situationsparametern und Erwartung auf sein Weltwissen zurückgreifen muss, um seine Entscheidungen und Handlungen dahingehend anzupassen.

Die Bedeutung von Situationsbewusstsein für die Luftfahrt ist deshalb so zentral, da es ein internalisiertes mentales Modell des momentan vorhandenen Umweltzustandes beschreibt (Endsley, 1999), der sich plötzlich oder schleichend verändern kann und auf den ein Pilot sein Denken und Handeln entsprechend einstellen muss, um geeignete Entscheidungen zu treffen. Auf Grund der ho-

hen Informationsdichte ist es notwendig, dass der Pilot beim Treffen seiner Entscheidungen und seiner Anpassung an die sich verändernde Umwelt durch ein MMS-Design gestützt wird, welches die verfügbaren Informationen integriert und an die Anforderungen des Piloten in der momentanen Situation anpasst (Endsley, 2001). Das heißt, dass die Informationen der MMS Cockpit, die den Piloten über den augenblicklichen Umweltzustand informieren, in einer Form gestaltet werden müssen, dass ein effektives und effizientes Arbeiten ermöglicht wird. Für das Design der Anwendung bedeutet das, dass die notwendigen Informationen in der richtigen Art und Weise und zum richtigen Zeitpunkt präsent sein müssen. Das heißt, dass kritische Hinweisreize geeignet bestimmt und dem Piloten präsentiert werden müssen, dass die Erwartungen, die der Nutzer an die MMS stellt, erfüllt werden, weiterhin die Aufmerksamkeit des Piloten gezielt mittels Design gerichtet wird und dass das System im Allgemeinen auf die Erfüllung der Pilotenziele ausgerichtet ist (Endsley, 1999). Geschieht dies nicht, kann performantes und effizientes Handeln in einem komplexen sozio-technischen System wie einem Flugzeug gefährdet sein. So stellt Wickens (2003) folgende Prinzipien der Displaygestaltung auf, um die Informationen, die der Pilot im Cockpit nutzen soll, in geeigneter Form zu präsentieren.

- *Principle of Information Need*
Es werden nur Informationen bereitgestellt, die der Pilot zur Aufgabenerfüllung benötigt
- *Principle of Legibility*
Die bereitgestellten Informationen sind interpretierbar, lesbar und werden unabhängig von der Modalität in geeigneter Form präsentiert
- *Principle of Display Integration/ Proximity Compatibility Principle*
Informationen, die zusammengehören, werden in räumlicher Nähe angezeigt
- *Principle of Pictorial Realism*
Die bildliche Repräsentation der Information sollte den realen Gegebenheiten nahekommen
- *Principle of the Moving Part*
Objektbewegungen auf dem Display entsprechen dem mentalen Modell des Piloten
- *Principle of Predictive Aiding*
Prognostizierende Informationen sind in einer Form darzustellen, die dem Piloten nutzt
- *Principle of Discriminality*
Informationen, die unterschiedlich sind, sind unterscheidbar darzustellen

Werden diese Prinzipien bei der Auslegung der MMS Cockpit beachtet, wird Entwicklung und Erhalt des Situationsbewusstseins des Piloten gestützt, was wiederum ein effizientes und effektives Arbeiten zur Erreichung der Nutzerziele ermöglicht. Es folgt hieraus die Bedeutung von Situationsbewusstsein für die Cockpitgestaltung und weshalb dieses als Kriterium zur Erhebung der MMS erfasst werden muss.

3.2 Workload - Beanspruchung und Arbeitsbelastung

Die ansteigende Komplexität im Cockpit sollte auch bei zukünftigen, neuen oder geänderten Cockpit-Auslegungen beachtet werden, um den Piloten nicht mit zu vielen Informationen zu belasten, die dieser in der ihm gegebenen Zeit nicht verarbeiten kann (Wierwille, Williges & Schiflett, 1979). Während sich Situationsbewusstsein primär auf die inhaltlichen Aspekte der Informationsverarbeitung in einer Situation bezieht (Rabl, 2016), beschreibt Workload das Ausmaß an Kapazität, die einem Operateur verbleibt, um eine Aufgabe zu bearbeiten (O'Donnell & Eggemeier, 1986). Einen breiteren Versuch, den Begriff Workload zu definieren, zitieren Hart und Wickens (1990), welcher auf das Aufmerksamkeits-Modell von Kahneman (1973) zurückzuführen ist. In dieser Definition liegt der Fokus folglich auf der Bindung von Ressourcen wie Aufmerksamkeit, Kognition oder Antwortkapazität in Form von den Kosten, die ein Mensch aufbringt, um den Anforderungen einer Aufgabe an der MMS gerecht zu werden. Jedoch bilden verschiedene Definitionen das theoretische Konstrukt Workload ab (Wierwille et al., 1979) und daher umfasst dieser Begriff neben der Ressourcenbindung als interne Nutzeranforderung auch die externalen Aufgabenanforderungen sowie die Nutzerleistung (Gartner & Murphy, 1976). Unter den Anforderungen einer Aufgabe, dem ‚Taskload‘, werden die Ziele verstanden, die notwendig sind, um die Aufgabe erfolgreich, in der gegebenen Zeit und mit einem bestimmten Leistungslevel zu erreichen (Gawron, 2008). Workload wird somit sowohl durch externe Faktoren wie Aufgaben-Schwierigkeit, Aufgabepriorität und situationsbedingte Abhängigkeiten bestimmt wie auch durch personen-spezifische Faktoren wie die Aufmerksamkeits- und Verarbeitungsressourcen des Nutzers. Beide Aspekte werden jedoch durch die Fähigkeiten und Erfahrungen des agierenden Menschen beeinflusst (Vidulich & Tsang, 2012). Auch Hart und Staveland (1988) nehmen diese Unterscheidung vor und fassen zusammen, dass die subjektiv vom Operateur empfundene Workload durch verschiedene, zusätzliche Faktoren, die in Form äußerlicher Gegebenheiten hinzukommen, wie sie beispielsweise durch die Anforderungen der Aufgabe entstehen, beeinflusst wird. Sie verorten daher die Workload auf die menschliche Seite der MMS und betonen ebenso die Kosten, die dieser aufbringen muss, um einen bestimmten Grad an Leistung zu erzielen. Auf Grund der Vielfalt an Definitionen schlägt Megaw (2005) ein Rahmenkonzept für Workload-Messungen vor, welches sich zwischen den Aufgabenanforderungen und der Aufgaben-Schwierigkeit, den Nutzeranforderungen und der Nutzerleistung als Hauptkomponenten und weiteren Stressoren aufspannt und sich auf die Interaktionen zwischen diesen Hauptkomponenten bezieht.

Das Konstrukt der Workload wird von dem der mentalen Ermüdung abgegrenzt. Gartner und Murphy (1976) beleuchten in diesem Zusammenhang mehrere Formen der Ermüdung:

- Ermüdung als ein Gefühl der Abgeschlagenheit
- Ermüdung als klinisches Syndrom
- Ermüdung als Ursache von Leistungs- und Fertigkeitseinbußen
- Ermüdung als neurophysiologischer Zustand
- Ermüdung im Sinne von Aufwendung von Energie

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

Zusammenfassend versteht man unter mentaler Ermüdung laut der Norm EN ISO 10075-1 eine zeitlich begrenzte mentale und/oder physische Beeinträchtigung der Effizienz als Folge und auf Grund mentaler Beanspruchung, die durch Erholung wieder gelöst werden kann. Somit ist Ermüdung vielmehr als Folge einer übermäßigen Arbeitsbelastung zu sehen und daher von dem Konstrukt der Workload zu unterscheiden.

Für die Entwicklung und das Design der MMS ist Workload unter Beachtung des Aufmerksamkeitsmodells nach Kahneman (1973) deshalb von großer Bedeutung, weil die Auslastung des Piloten bei der Bedienung der MMS dahingehend beachtet werden muss, ob dieser in der Lage ist, weitere, zusätzliche Aufgaben zu bearbeiten, die Komplexität vorhandener Aufgaben zu durchdringen und ob es dem Piloten möglich ist, für weitere, plötzlich auftretende Ereignisse genug Kapazitäten aufzubringen, um diesen zu begegnen (Wickens, Hollands, Banbury & Parasuraman, 2013). Die verfügbaren Ressourcen des Piloten und deren Verbrauch sind somit für die Gestaltung der MMS zentral, da diese auch den Leistungsgrad des Piloten mitbestimmen (Wickens et al., 2013). So kann laut Wickens et al. (2013) einem übermäßiger Verbrauch an Ressourcen in einer ‚Overload‘-Situation wie sie beispielsweise in komplexen Situationen mit mehreren parallel zu bearbeiteten Aufgaben stattfindet mit Hilfe folgender Methoden begegnet werden:

- Reduktion der Belastung des Arbeitsgedächtnisses
- Neuorganisation von Arbeitsweisen
- Delegation an weitere Operateure
- Automatisierung

Diese Maßnahmen sind in der Designentwicklung von Cockpit-Modifikationen zu berücksichtigen. Eine Reduktion der Workload und somit auch der Belastungen des Arbeitsgedächtnisses wird beispielsweise durch eine geeignete Darstellung von Informationen auf den Displays (respektive den Multifunktionsdisplays) möglich (Kantowitz & Casper, 1988), was ebenso zu einer Neuorganisation vorhandener Arbeitsweisen führen kann. Neben der Verteilung von Aufgaben an weitere Operateure kann, wo dies aus verschiedenen Gründen nicht möglich ist und die Arbeitsbelastung hoch ist, auch auf die Automatisierung von Aufgaben zurückgegriffen werden. Automatisierung kann zu einer Reduktion der Workload führen, sofern diese in geeigneter Form in das Design überführt wird (Wickens et al., 2013), aber auch das Leistungsvermögen des Gesamt-Systems in taktischen Situationen verbessern (Wilkinson, 1992). Da im Falle einer Automatisierung jedoch Aufgaben vom technischen System beziehungsweise der Maschine übernommen werden, obliegt es dem Piloten, diese Ausführung der Aufgaben zu überwachen, was wiederum Workload erzeugt (Kantowitz & Casper, 1988). Im Extremfall kann eine Automatisierung zu einer Erhöhung der Workload im Vergleich zu teilautomatisierten Systemen führen, was Bainbridge (1983) als Paradoxie der Automatisierung bezeichnet. Um diese Interaktion daher nicht zu gefährden, ist es notwendig, dass die Fähigkeiten und die Effizienz des Nutzers gewahrt bleiben, indem der Nutzer adäquat in die Aufgabe und den momentanen Zustand des Systems eingebunden wird (Bainbridge, 1983). Um dies und somit eine Reduktion der Workload zu ermöglichen, ist es notwendig, das System in einer Art und Weise zu desig-

nen, dass dieses System auch mögliche Fehlentwicklungen berücksichtigt und mit dem Nutzer kontinuierlich und in geeigneter Form interagiert und ihn mit relevanten Feedback-Informationen versorgt (Norman, 1990).

3.3 Gebrauchstauglichkeit

Da die Interaktionen des Nutzers mit der Maschine bzw. dem System im Cockpit stattfinden, wird auch im Cockpit anhand der Bedienung offenbar, wie nutzbar die neue technische Funktion oder die Cockpit-Modifikation in der Anwendung ist. Als übergeordnetes Konstrukt oder Qualitätskriterium zur Bewertung von Cockpit-Modifikationen kann somit das Konstrukt der Gebrauchstauglichkeit gesehen werden. Die Gebrauchstauglichkeit ist definiert durch die DIN Norm (DIN EN ISO 9241-210) als:

„Ausmaß, in dem ein System, ein Produkt oder eine Dienstleistung durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um festgelegte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“

In diesem Zusammenhang wird die Gebrauchstauglichkeit durch die objektivierbaren Maße Effektivität und Effizienz sowie durch das subjektive Maß Zufriedenheit bestimmt. Effektivität ist hierbei definiert als die Genauigkeit und Vollständigkeit, mit der ein Arbeitsergebnis erreicht werden soll. Unter Effizienz wiederum wird das Verhältnis der eingesetzten Ressourcen wie beispielsweise Zeit, zu dem erreichten Grad an Vollständigkeit verstanden. Das letzte Kriterium zur Bestimmung der Gebrauchstauglichkeit ist die Nutzerzufriedenheit, welches als rein subjektives Maß zu verstehen ist (DIN EN ISO 9241-210).

Die drei Maße tragen im Wesentlichen dazu bei, wie nutzbar das Produkt für den Nutzer ist, um die angestrebten Ziele im Nutzungskontext zu erreichen. Hieraus abzuleiten ist erneut das Erreichen des Nutzerziels, da sich hieran auch ein Vorgehen hin zum nutzerzentrierten Design orientieren sollte – also die zentrale Frage, was konkret die Nutzeroberfläche des zu entwickelnden Produkts/ der Funktion dazu beitragen kann, damit der Nutzer seine Ziele, Aufgaben und Bedürfnisse erfüllen kann (Endsley & Jones, 2012). Hierzu sollten in der Entwicklung und dem Design von technischen Funktionen und Displayinhalten einige Prinzipien eingehalten werden wie sie auch Karwowski (2001) beschreibt:

- *Consistency*
Ähnliche Aufgaben sollten in analoger Art und Weise bearbeitet werden. Eingaben und Ausgaben – also die Dialoge zwischen Nutzer und System – sollten innerhalb der Menüs und Aufgaben gleich gehalten werden (Williges, Williges & Fainter, 1988).
- *Compatibility*
Die Handhabung einer Schnittstelle sollte in einer Form umgesetzt werden, die für den Nutzer auf Grund seines Wissens und seiner Interaktion mit anderen Produkten sowie auf Grundlage von Konventionen erwartbar ist (Ravden & Johnson, 1989).

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

- *Consideration of user resources*
Bei der Gestaltung sollte der Ressourcen-Einsatz des Nutzers mitberücksichtigt werden. Insbesondere Belastungen des Kurzzeitgedächtnisses sollten nach Möglichkeit minimiert werden, indem die Displayinhalte möglichst einfach und verständlich gestaltet werden (Shneiderman et al., 2005).
- *Feedback*
Das System/ Produkt sollte dem Nutzer Rückmeldung hinsichtlich dessen Eingaben geben und ihn über die Ergebnisse seines Handelns informieren. Diese können in Abhängigkeit der Häufigkeit der Nutzung und deren Bedeutsamkeit unterschiedlich salient sein (Shneiderman et al., 2005).
- *Error prevention and recovery*
Fehleingaben durch den Nutzer sollen anhand des Produktdesigns minimiert werden. Es sollten Möglichkeiten für den Nutzer bestehen, die Eingaben zu überprüfen und diese zu verifizieren. Fehlerhafte Eingaben durch den Nutzer sollten leicht, schnell und einfach rückgängig gemacht werden können (Ravden & Johnson, 1989).
- *User control*
Der Nutzer sollte hinsichtlich seiner Handlungen ein größtmögliches Maß an Kontrolle hinsichtlich seiner Eingaben und der resultierenden Rückmeldungen ausüben können. Überraschendes Systemverhalten ist zu vermeiden (Shneiderman et al., 2005).
- *Visual clarity*
Die dargestellten Informationen müssen gut und leicht lesbar sein ohne zu Missverständnissen zu führen (Karwowski, 2001).
- *Prioritisation of functionality and information*
Wichtige Informationen müssen besonders leicht zugänglich sein (Karwowski, 2001).
- *Appropriate transfer of technology*
Auch Technologien anderer Produkte sollten verwendet und in geeignetem Maß integriert werden, sofern diese die Gebrauchstauglichkeit des zu entwickelnden Produktes verbessern (Karwowski, 2001).
- *Explicitness*
Die zur Verfügung gestellten Hinweise wie Icons oder Analogien sollten hinsichtlich ihrer Funktion und ihres Vorgehens eindeutig sein. Anhand dieses Prinzips kann gewährleistet werden, dass der Nutzer ein eindeutiges und klares Verständnis hinsichtlich des Systems entwickelt (Ravden & Johnson, 1989).

Ein weiteres Prinzip der Gestaltung von interaktiven Schnittstellen ist nach der Norm DIN EN ISO 9241-210 die Erlernbarkeit des Systems. Diese Eigenschaft bezieht sich demnach auf die Geschwindigkeit, mit der ein Nutzer auch komplexe Systeme erlernen kann, zu bedienen (Wickens, Gordon & Liu, 1998). Der Nutzer soll daher Eingabemöglichkeiten und Aktivitäten des Systems sowie dessen Bedienung möglichst schnell verstehen und somit erlernen können.

Da das Cockpit hochagiler Luftfahrzeuge in erster Linie für den eigentlichen Einsatz ausgelegt sein muss und der Pilot dahingehend unterstützt werden muss, variiert das Ausmaß an Bedeutsamkeit der obengenannten Prinzipien dahingehend, dass ein Design frühestens dann als gebrauchstauglich anzusehen ist, sofern dieses nicht schlechter als die bisherige Cockpit-Auslegung ist (J. Turner, 1997). Diese pessimistische Sichtweise kann allerdings nicht als Maßstab für die Weiterentwicklung bisheriger und zukünftiger Systeme verwendet werden, da sie insbesondere den subjektiven Aspekt der Gebrauchstauglichkeit – die Nutzerzufriedenheit – gänzlich auslässt. Für die Umsetzung der Usability-Prinzipien in das Cockpit eines Luftfahrzeuges sind folglich bei der Implementierung von weiteren technischen Funktionen oder Cockpit-Modifikationen drei Betrachtungsebenen nach Williges et al. (1988) zu berücksichtigen. So werden die Rohdaten von Sensoren und Systemen vorverarbeitet und in mehreren aufeinander aufbauenden Softwareschichten dem Nutzer präsentiert. Der Entwicklungsprozess einer Cockpit-Modifikation ist mehrstufig und notwendigerweise iterativ und entsprechend müssen bei der Entwicklung die Prozesse und Verfahren der vorhandenen Software beachtet werden (Williges et al., 1988). Im Rahmen dieser drei Betrachtungsebenen sind die oben genannten Design-Prinzipien als bestmögliches Vorgehen zu berücksichtigen. Zur Bewertung und Evaluierung der Gebrauchstauglichkeit sind jedoch Kriterien notwendig, die aus dem Nutzungskontext hervorgehen, da diese die spezifischen Erwartungen an das System beschreiben können (Banks, Plant & Stanton, 2018).

3.4 Das multidimensionale Erhebungskonzept zur Bewertung von MMS

Die Qualität der Nutzerschnittstelle Cockpit soll anhand der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit beurteilt werden. Es ist jeweils festzustellen, ob und in welchem Ausmaß eine Änderung des Cockpits durch Implementierung einer technischen Funktion das Aufrechterhalten des Situationsbewusstseins des Piloten beeinträchtigt, zu einer übermäßigen Workload führt und inwiefern die Nutzbarkeit der MMS weiterhin sichergestellt ist. Hierzu ist es notwendig, die technische Funktion mit Hilfe von Piloten als Experten hinsichtlich dieser Kriterien und mit Hilfe von hierfür geeigneten Methoden zu evaluieren und abschließend zu bewerten.

Zentral für das Systemverständnis ist hier die Rolle des eigentlichen Nutzers als Experte und Spezialist für sein Arbeitsfeld. Dieser ist nicht einfach nur als Studienteilnehmer zu sehen, sondern auch als Teil des Systems, der ebenso ein Interesse hat, dieses zu verbessern (Shorrock, Leonhardt, Licu & Peters, 2014). Seine Rolle in der Systementwicklung ist es daher nicht nur, ein Nutzungsbedürfnis zu äußern, sondern auch die Ereignisse seines Arbeitsumfeldes zu beleuchten, zu analysieren und zu schlussfolgern, um zur Verbesserung des Gesamt-Systems in jeder dieser Ebenen beizutragen. Die Einbindung des Nutzers tritt mit der weiteren Ausreifung des Produktes immer weiter zurück; doch ist die Integration des Nutzers in den Gesamtentwicklungsprozess und daher auch in die Evaluierung der Cockpit-Modifikation von entscheidender Bedeutung, weil dies zur Validität und Nutzbarkeit der Datensammlung und der hieraus abgeleiteten Verbesserungen beiträgt (Shorrock et al., 2014).

Die zu verwendenden Methoden, die in erster Linie dazu geeignet sein müssen, das ihnen zu eigene Konstrukt zu erfassen, sollen folglich des Weiteren dazu geeignet sein, das Wissen und die An-

sichten der Experten zielgerichtet zu erheben und mit objektiven Messwerten kontrastiert werden. Das setzt die Einbindung der Nutzer in den Entwicklungs- und Designprozess jedoch in strukturierter Weise voraus, da lediglich verbal geäußerte Meinungen von Experten ohne experimentelle Rückkopplung zu volatilen Ergebnissen führen können. Stattdessen sollten die subjektiven Expertenmeinungen genutzt werden und in objektiv messbare Kriterien und Konstrukte überführt werden (Perrot, Schader, Leonhardt & Licu, 2019).

Perrot et al. (2019) beschreiben für die Luftverkehrskontrolle das Problem, dass allgemeine Ansätze an Human-Factors-Methoden wie Interviews, Beobachtungen und Befragungen in den Rahmenbedingungen von nutzerzentriertem Design nicht strukturiert eingesetzt werden, um den Nutzer zielgerichtet einzubinden. Daher ist es notwendig, die zu erfassenden Kriterien an dem zu designenden Objekt auszurichten. Um sowohl subjektive als auch objektive Bewertungskriterien sinnvoll für die drei Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit zu erheben, ist daher ein Konvolut aus mehreren subjektiven und objektiven Methoden notwendig, um den Nutzer optimal in den Designprozess zu integrieren.

Zur Bestimmung der Qualität der Mensch-Maschine-Schnittstelle Cockpit anhand der Qualitätskriterien ist deshalb ein multidimensionales Methoden-Inventar zu entwickeln. Es sind repräsentative Methoden auszuwählen, welche sich an die Simulationsumgebung schnell und adäquat anpassen lassen. Die Methoden sollten daher flexibel in ihrer Anwendung sowie in der Erhebung der Qualitätskriterien validiert sein. Zudem sollten die eingesetzten Methoden die spezifischen Aspekte der Cockpit-Modifikationen erfassen, welche für deren Design und einem möglichen Re-Design relevant sind. Das heißt, die Methoden des Methoden-Inventars sollen sich gut zur Bestimmung der Cockpit-Modifikation in den situativen Kontext einbetten, in welchem sie zu messbaren Ergebnissen führen sollen. Sie können des Weiteren kombiniert miteinander verwendet zur Messung der Qualitätskriterien eingesetzt werden.

3.4.1 Methodenauswahl zur Erfassung von Situationsbewusstsein

Die Evaluierung von Situationsbewusstsein bezieht sich auf die Erhebung des Wissens, das für die jeweils momentan durchgeführte Aufgabe des Nutzers relevant ist (Gawron, 2008). Situationsbewusstsein bezieht sich demnach auf das Wissen über die dynamischen Entwicklungen der Umwelt, welches der Nutzer durch Wahrnehmungsprozesse erlangt und in Relation zu seinen Nutzerzielen setzt und hinsichtlich deren zukünftiger Entwicklungen beurteilt, bevor er eine Entscheidung trifft und danach handelt (Endsley, 1995).

Nach Fracker (1991) können die Messmethoden des Situationsbewusstseins in drei unterschiedliche Strategien unterteilt werden: Explizite und implizite Messverfahren sowie subjektive Messverfahren.

Eine explizite Messstrategie kann als direktes Vorgehen zur Erhebung von Situationsbewusstsein verstanden werden, da von dem Nutzer bewusste Aspekte erfragt werden. Unterschieden wird in der Strategie der expliziten Erhebung nach retrospektiven Verfahren situativer Ereignisse und der Erhe-

bung von konkurrierender Gedächtnisabfragen wie sie beispielsweise während einer Simulation vorgenommen werden können. Der Unterschied dieser beiden Vorgehen ist darin ersichtlich, dass bei der expliziten retrospektiven Erhebung die Erinnerung des Probanden zu situativ relevanten Inhalten mit einem zeitlichen Versatz erhoben wird, während bei der expliziten Erhebung von konkurrierenden Gedächtnisinhalten die Abfrage direkt in der Situation stattfindet (s.a. Rabl, 2016). An dieser Stelle wird wiederum die Diskussion um das Situationsbewusstsein als Produkt und Prozess offenbar, da hier Entscheidungen hinsichtlich der zu verwendenden Methoden und deren Befragungstechniken zu treffen sind (Durso & Sethumadhavan, 2008). Produkt-orientierte Befragungen zielen hier auf reine Gedächtnisabfragen ab, während am Prozess orientierte Methoden weitere Aspekte wie Leistungsparameter mit in Bezug zu situationsabhängigen Befragungen setzen, um die zweite Ebene des Situationsbewusstseins ‚Verständnis‘ näher zu beleuchten (Durso & Sethumadhavan, 2008).

Implizite Messvorgehen hingegen beschreiben das Situationsbewusstsein indirekt anhand von objektiven Leistungsmessungen als Indikator für vorhandenes Situationsbewusstsein. Im Rahmen dieser Strategie ist es zwar möglich, objektive Daten zu erheben; diese bergen jedoch die Gefahr, durch ein anderes Konstrukt beeinflusst zu werden als durch das Situationsbewusstsein (Uhlarik & Comerford, 2002).

Die einfachste Form Situationsbewusstsein zu erheben, ist die Anwendung von subjektiven Messmethoden (Pew, 2000). Unter einer subjektiven Messstrategie versteht man die subjektive und retrospektive Selbst- oder Fremdeinschätzung des Situationsbewusstseins (Uhlarik & Comerford, 2002) anhand von numerischen Werten über ein bestimmtes Aufgabensegment oder im Hinblick einer Mission (Rabl, 2016). Nach Fracker (1991) kann dieses Vorgehen entweder direkt oder vergleichend erfolgen; das heißt entweder in Bezug auf eine bestimmte Aufgabe oder anhand von mehreren Aufgaben, in denen das subjektiv empfundene Situationsbewusstsein erhoben wird. Subjektive Messungen haben den Vorteil, dass diese in der Regel kostengünstig und leicht anzuwenden sind. Sie stören des Weiteren nicht das Simulationsgeschehen und die Nutzeraufgabe, sind jedoch auf Grund ihres retrospektiven Charakters anfällig für Gedächtnis-Verzerrungen. Allerdings sind viele subjektive Messungen nicht ausreichend validiert und spiegeln nicht unbedingt das Situationsbewusstsein des Nutzers wider und sollten daher möglichst nur in Verbindung mit weiteren objektiven Messungen genutzt werden (Jones, 2000).

Den genannten Messstrategien nach Fracker können jeweils verschiedene Kategorien an Messmethoden zugeordnet werden wie sie auch Rauch (2009) identifiziert. Demnach können die Messmethoden in Form von Prozessmaßen, Rating-Skalen, Befragungsverfahren oder anhand von Verhaltensmaßen erhoben werden. Unter Prozessmaßen versteht man hier kontinuierliche Erhebungen wie psycho-physiologische Messungen, aber auch verbale Protokolle im Sinne von Think-Aloud-Techniken, während Ratingverfahren insbesondere die subjektive Messstrategie beschreiben. Unter Beobachtungsverfahren werden hingegen objektive Messungen verstanden, wie sie durch das Abfragen von Gedächtnisinhalten im situativen Kontext entstehen. Abschließend nennt Rauch (2009) Verhaltensmaße als Verbund von Messungen, die Leistungsparameter erheben, also Vertreter der impliziten Messstrategie sind.

3.4.1.1 *Situation Awareness Rating Technique - SART*

Der SART als Messvorgehen ist ein weitverbreiteter subjektiver Fragebogen zur Erhebung von Situationsbewusstsein (Endsley, Selcon, Hardiman & Croft, 1998; Jones, 2000; Salmon, Stanton, Walker et al., 2009) nach Taylor (1990) und ein Vertreter einer subjektiven und retrospektiven Messstrategie zur Bestimmung des Situationsbewusstseins. Ursprünglich ist dieser Fragebogen auf eine Befragung von Piloten der Royal Airforce zurückzuführen und wurde als Evaluierungswerkzeug für Cockpit und Besatzungssysteme entwickelt (Selcon & Taylor, 1996). Die Piloten wurden in drei-stufigen Interviews zu Szenarien, Konstrukt-Eingrenzung und Validierung der Konstrukt-Struktur hinsichtlich ihres Situationsbewusstseins befragt. Aus den Ergebnissen konnten zehn Faktoren (10D SART) abgeleitet werden, die des Weiteren auf drei Faktoren (3D SART) reduziert werden können. Eine Bewertung erfolgt anhand einer siebenstufigen Likert-Skala mit den Ankern 1 (niedrig) und 7 (hoch) (Selcon & Taylor, 1996; Taylor, 1990). Die Skalen lassen sich daher direkt auf die Aspekte und das Wissen zurückführen, welche die Piloten als für ihr Situationsbewusstsein relevant erachteten (Jones, 2000) und welche auf der verwendeten und abgestimmten Arbeitsdefinition von Situationsbewusstsein basieren, die Taylor (1990) vorstellt:

„Situational awareness is the knowledge, cognition and anticipation of events, factors and variables affecting the safe expedient and effective conduct of the mission.“

Maßgeblich für das Situationsbewusstsein sind demnach die zehn Faktoren: Instabilität der Situation (1), Variabilität der Situation (2), Komplexität der Situation (3), Erregungsgrad (4), freie mentale Kapazität (5), Konzentration (6), Aufmerksamkeitsverteilung (7), Informationsmenge (8), Informationsqualität (9) und Vertrautheit (10). Diese Aspekte des Situationsbewusstseins können in die drei Gruppierungen des 3D SARTs – in Beanspruchung der Aufmerksamkeits-Ressourcen (1-3), Verfügbarkeit an Aufmerksamkeits-Ressourcen (4-7) und Verständnis der Situation (8-10) subsumiert werden (Selcon & Taylor, 1996; Taylor, 1990). Weiterhin kann anhand der Skalenwerte des 3D SART ein übergeordneter Gesamtwert des Situationsbewusstseins errechnet werden, indem man die Differenz des Skalenwerts ‚Verständnis der Situation‘ und der Differenz der Skalenwerte ‚Beanspruchung von Aufmerksamkeits-Ressourcen‘ und ‚Verfügbarkeit von Aufmerksamkeits-Ressourcen‘ (Situationsbewusstsein = Verständnis – (Beanspruchung – Verfügbarkeit) bildet (s.a. Jones, 2000; Stanton et al., 2011).

Üblicherweise wird dieses Vorgehen zur Selbsteinschätzung des Situationsbewusstseins nach Durchführung der Nutzeraufgabe eingesetzt. Die Anwendung dieses Verfahrens hat den Vorteil, dass diese Methodik leicht und schnell bei geringem Ressourcen-Einsatz zu verwenden ist, ohne die Nutzeraufgabe zu beeinflussen. Von Nachteil ist allerdings, dass bei subjektiven Selbsteinschätzungen wie dem SART, der Nutzer sich seiner Selbsteinschätzung hinsichtlich seines Situationsbewusstseins ‚bewusst‘ sein muss und sich an diesen Zustand im Nachhinein erinnern muss. Das führt mitunter dazu, dass Korrelationen zu Nutzerleistung entstehen und dies ebenso die Sensitivität dieser Methodik beeinflussen kann (Salmon, Stanton, Jenkins & Walker, 2009). Bereits Selcon und Taylor (1996) folgern, dass der SART als Methodik in Zusammenhang mit der Nutzerleistung steht. Sie stellen fest,

dass eine Verbesserung der Dimension ‚Verständnis der Situation‘ mit einer schnelleren Nutzerleistung einhergeht und ziehen daraus den Schluss, dass der SART diagnostisch ist. Die Tatsache, dass Situationsbewusstsein in mehreren Dimensionen gemessen wird, liegt daran, dass der Autor des SART Situationsbewusstsein als komplexes Konstrukt ähnlich wie Aufmerksamkeit und Kognition (Taylor, 1990) oder Workload (Gawron, 2008) benennt. Zumindest ist aus diesem Grund auch naheliegend, weshalb Selcon, Taylor und Koritsas (1991) feststellen, dass der SART ebenfalls durch eine Veränderung der Aufgaben-Schwierigkeit moderiert wird, jedoch sensitiv auf die Erfahrung der Probanden reagiert. Zusammenfassend ist der SART eher als globales Messinstrument zu sehen, das weniger aufgabenspezifisch einzusetzen ist, als vielmehr bezüglich der Selbsteinschätzung der Probanden hinsichtlich ihres Situationsbewusstseins, ihres Wissens und ihrer Leistung (Salmon, Stanton, Walker et al., 2009). Ergänzend sind die Ergebnisse von Endsley et al. (1998) zu sehen, in denen nahegelegt wird, dass die Einschätzungen der Nutzer mittels SART mit deren Einschätzungen bezüglich ihres Selbstvertrauens korrelieren und daher anhand des SART nur deren wahrgenommenes Situationsbewusstsein gemessen wird.

Für die Verwendung des SART im multidimensionalen Methoden-Inventar spricht aber neben den offensichtlichen Vorteilen wie schneller und einfacher Anwendung bei geringem Auswerteaufwand, dass dieses Messinstrument global einsetzbar ist und explizit für die Erhebung von Situationsbewusstsein der Besatzung von hochagiler Luftfahrzeuge entwickelt wurde und daher einen hohen Grad an ökologischer Validität besitzt (Jones, 2000).

3.4.1.2 *Situation Awareness Global Assessment Technique -SAGAT*

Der SAGAT als Messmethodik ist ein explizites Messverfahren konkurrierender Gedächtnisinhalte, das von Endsley (1988b) mit Luftfahrtbesatzungen aus dem operativen Kontext entwickelt und vorgestellt wurde. Bevor eine Anwendung des SAGAT erfolgen kann, ist die Durchführung einer Bedarfsanalyse notwendig, wie sie Endsley (1993) in Form einer Situation Awareness Requirements Analyse (SARA) umsetzt. Dies ist ein Vorgehen, das es erlaubt, die dynamischen Situations-Parameter und deren Bedeutung für die Akteure in einer zu fliegenden Mission mit Hilfe von Befragungen und einer hierauf basierenden zielgerichteten Aufgabenanalyse zu bestimmen. Ziel der Anwendung einer SARA ist es, den jeweiligen Informationsbedarf der Operateure gemäß der drei möglichen Ebenen des Situationsbewusstseins – Wahrnehmung, Verständnis und Projektion – auf zukünftige Ereignisse festzuhalten. Die Anwendung des SAGATs erfolgt daher auf Grundlage der durchgeführten SARA, weil hier das Wissen identifiziert wird, das später im Vorgehen eines SAGATs erfragt werden kann (Endsley, 2000a). Eine Umsetzung des SAGATs erfolgt unter der Voraussetzung, dass ein Simulator und ein zu testender Prototyp der Designänderung vorhanden sind. Hierin ist es Aufgabe des Nutzers, eine Mission oder ein hierin integriertes Szenario zu fliegen, welches an geeigneten Stellen angehalten wird. Das Szenario wird zu randomisierten Zeitpunkten unterbrochen und die Displayinhalte ausgeblendet oder verdeckt. Zu diesen Zeitpunkten wird der Nutzer zu mehreren, in der SARA ermittelten, situativen Informationsanforderungen befragt, die nach Möglichkeit dynamisch sein sollen. Weitere Befragungen des Situationsbewusstseins finden dann zu anderen Zeitpunkten statt. Die Fragestellungen sind hierbei verschieden, um zu verhindern, dass der Pilot ein spezifisches Augenmerk auf die zu er-

fragenden Informationen lenkt. Nach Abschluss der Befragungen werden die Antworten des Piloten evaluiert und mit den Simulationsdaten zur Bildung eines Gesamtwertes verglichen (Endsley, 1990). Die Objektivität des Messvorgehens leitet sich laut Endsley (1988b) aus dieser Gegenüberstellung von Simulationsdaten und der korrekten Beantwortung der Fragestellungen durch den Piloten ab – ein Vorgehen, das optional auch durch andere operative Experten vorgenommen werden kann.

Kritik an der Messung des Situationsbewusstseins anhand des SAGATs wird unter anderem durch Fracker (1988) geäußert, da dieses Vorgehen nur in Simulationen verwendet werden kann und mitunter sichergestellt werden muss, dass die Leistungen des Piloten anhand der Intrusivität des SAGATs beeinflusst werden können. Endsley (1988b; Endsley, 2000a) hingegen verneint einen Einfluss auf die Leistungen der Piloten durch die Durchführung einer Untersuchung mit und ohne SAGAT-Unterbrechungen.

Ein weiterer Kritikpunkt wird von Durso und Sethumadhavan (2008) geäußert. Die Autoren bezeichnen den SAGAT als eine Messmethode, die lediglich Gedächtnisabfragen vornimmt bzw. das Ergebnis eines Verständnisprozesses misst, nicht aber, wie der Operateur zu diesem Verständnis gelangt. Durso et al. (1998) entwickelten daher auch die Situation Awareness Present Assessment Method (SPAM), die Messungen von Situationsbewusstsein direkt im Szenario vornimmt, ohne dieses zu unterbrechen und mit einbezieht, dass die Informationssuche auch ein Teil des Situationsbewusstseins darstellt. Ein Messkriterium des SPAMs ist daher die Latenzzeit, die ein Operateur benötigt, um auf eine Abfrage von Situationsbewusstsein zu reagieren, anstatt der Abfrage von Gedächtnisinhalten.

Bei einem Vergleich der beiden Messvorgehen SPAM und SAGAT beschreiben Durso et al. (1998), dass sowohl SAGAT als auch SPAM in ihrer Untersuchung von Lotsen der Flugverkehrskontrolle die Aussagen von Expertenurteilen und verbleibenden Handlungen vorhersagen. Endsley, Sollenberger und Stein (2000) können in ihrer Untersuchung, in welcher verschiedene Displayformate im Bereich der Luftverkehrskontrolle eingeführt wurden, keinen Zusammenhang zwischen den gemessenen Latenzzeiten und Messungen des Situationsbewusstseins der ersten Ebene herstellen. Sie ergänzen weiterhin, dass die Befragungen während des Szenarios zu keiner Unterscheidung der Displayformate führen, während dies bei Verwendung des SAGATs der Fall war.

Endsley (1988b) rechtfertigt den Einsatz des SAGATs mitunter durch ihren kritischen Einwand, dass Situationsbewusstsein in Untersuchungen mit gängigen Methoden oft nur eine spezifische Designänderung analysieren und so Situationsbewusstsein nur eingeschränkt und fokussiert erfasst wird, weil die Aufmerksamkeit der Operateure auf diese spezifische Funktion gelenkt wird. Das Situationsbewusstsein könne so allerdings nicht zufriedenstellend erfasst werden, da Situationsbewusstsein ein globales Konstrukt ist.

Der SAGAT als Messvorgehen ist für die Erhebung von Besatzungen hochagiler Luftfahrzeuge entwickelt worden und daher für die Verwendung zur Überprüfung der Cockpit-Qualität nach Designänderungen geeignet, zumal die Datenerfassung laut Endsley (1988a) als objektiv und global anzusehen ist, einen Einblick in das augenblickliche Situationsverständnis und -Wissen des Piloten erlaubt und

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

inhaltsvalide ist. Zudem ist die Einbindung einer Methodik wie dem SAGAT dann angezeigt, wenn der Erhalt des Situationsbewusstseins als Designziel definiert wird, da so während des Design-Prozesses ermöglicht werden kann, dass dieses Ziel erreicht wird (Endsley, 2001).

3.4.1.3 *Eye-Tracking*

Wenn, während der Pilot im Cockpit die zu untersuchende Nutzeraufgabe erledigt, man die Eye-Mind-Hypothese nach Just und Carpenter (1980) zu Grunde legt, ist davon auszugehen, dass der Nutzer den Inhalt, den er gerade liest, zu verstehen versucht und dass der Nutzer mit dem Auge den Inhalt so lange fixiert, bis dieser verarbeitet ist. Diese Annahme ist für die Verwendung des Eye-Trackings als Methodik zentral, da anhand dieser Annahme festgelegt wird, dass sich der Nutzer die visuell erfassten Informationen so zu eigen macht. Eye-Tracking als indirektes Maß zur Erhebung von Situationsbewusstsein wurde bereits in verschiedenen Human-Factors-Bereichen der Luftfahrt erfolgreich angewandt. Van Dijk, van de Merwe und Zon (2011) konnten in einer Studie zur Untersuchung einer Fehlfunktion, die während des Fluges auftrat, aufzeigen, dass Verweildauer auf relevanten Displayinhalten sowie ein breitflächiges Überprüfen der Inhalte zu einem verbesserten Situationsbewusstsein führen. Dennoch betonen auch hier die Autoren die Wichtigkeit, Situationsbewusstsein mit mehreren Messmethoden zu erheben, um so ein kohärenteres Bild des Piloten und seines Situationsbewusstseins zu erheben. In einer Untersuchung mit Fahranfängern konnten Pradhan, Fisher und Pollatsek (2006) zeigen, dass nach einem Training die Fahranfänger ein vermehrtes Augenmerk auf risikorelevante Inhalte legten. In einem weiteren Experiment im Automobil-Kontext konnte durch Moore (2009) festgestellt werden, dass Situationsbewusstsein mit einer längeren Fixationsdauer und einer geringeren Fixationsanzahl auf relevante Displayinhalte einhergeht, sofern keine übermäßige Belastung der Probanden besteht. Aufbauend hierauf stellte die Autorin in einem Experiment zur Luftverkehrskontrolle fest, dass eine höhere Anzahl an Fixationen mit einer geringeren Anzahl an Fehlern einhergeht und wiederum die Fixationsdauer ein Indiz für Situationsbewusstsein darstellt (Moore, 2009). Eye-Tracking Messungen werden in der Luftfahrt weiterhin zur Unterscheidung unterschiedlicher Blickstrategien von Experten und Novizen genutzt, um beispielsweise die Blickstrategien von Novizen zu trainieren (Muehlethaler & Knecht, 2016; s.a. Robinski & Stein, 2013).

3.4.2 Methodenauswahl zur Erfassung von Workload

Die Workload der Nutzer kann direkt an der Mensch-Maschine-Schnittstelle (MMS) verankert werden, da diese in Abhängigkeit der Unterschiede von Operateuren und Systemen variiert (Hart & Wickens, 1990). Es sind daher Messstrategien zur Erhebung von Workload notwendig, die sich direkt in die Simulationsumgebung integrieren lassen, um die technischen Änderungen im Cockpit und deren Auswirkung auf die Workload der Piloten zu überprüfen. Relevant ist hier das Rahmenwerk nach Megaw (2005), der Aufgabenfaktoren, Nutzeranforderungen und Nutzerleistung sowie weitere, umweltbedingte Stressoren als beeinflussende Faktoren der Workload bezeichnet.

Einen Überblick über mögliche Messstrategien, die sich in dieses Rahmenkonzept integrieren lassen, geben unter anderem Wierwille et al. (1979), die Workload-Methoden nach folgenden Kategorien unterscheiden: Erhebung als subjektive Einschätzung, Ableitung der Workload aus Messungen in

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

der Hauptaufgabe, Workload-Erhebung als Ressourcenverbrauch im Sinne freier mentaler Kapazität und physiologische Messungen. Die Autoren unterteilen die subjektive Einschätzung in die Verwendung von Fragebögen, Interviews und nach psychometrischen Rating-Skalen.

Da die Termini Rating-Skalen und Fragebögen oft synonym verwendet werden, fügen sie hier an, dass nach ihrem Verständnis Fragebögen im Gegensatz zu Rating-Skalen nicht strikt an psychometrischen Skaleneigenschaften orientiert sind. Rating-Skalen zur Erhebung von Workload treten nach Hart und Wickens (1990) in drei unterschiedlichen Varianten auf – als ein- oder multidimensionale Rating-Skalen sowie als hierarchische Rating-Skalen. Unter einer eindimensionalen Ratingskala versteht man hier eine globale Messung, die Workload anhand einer einzigen Variablen erhebt. Multidimensionale Rating-Skalen beschreiben die Workload anhand mehrerer, einzeln zu bewertender Faktoren. Die Verwendung einer multidimensionalen Rating-Skala erlaubt einen detaillierteren Einblick in die empfundene Workload der Nutzer und sollte einen höheren Grad an Diagnostizität aufweisen, als dies bei einer eindimensionalen Skala der Fall ist, die zu einer Erstüberprüfung aufgrund ihrer Sensitivität ebenso zur Workload-Erhebung geeignet ist (Hill et al., 1992). Das Rating-Verfahren mit einer hierarchischen Rating-Skala basiert auf einem Entscheidungsbaum, der dem Nutzer zwei bis drei Alternativen zur Einschätzung der Workload zur Auswahl anbietet und in einem weiteren Entscheidungsverfahren spezifiziert. In der Anwendung können Rating-Skalen während der Aufgabe, nach Abschluss spezifischer Aufgabensegmente oder zum Abschluss der Nutzeraufgabe erfragt werden und entweder anhand einer Selbstauskunft oder mittels (Experten-) Beobachtern bewertet werden (Hart & Wickens, 1990). Subjektive Rating-Skalen sind als Methodik einfach, schnell und mit geringem Kostenaufwand anzuwenden, ohne sonderlich intrusiv zu sein (Stanton et al., 2011). O'Donnell und Eggemeier (1986) benennen andererseits fünf Einschränkungen, die im Zusammenhang mit der Verwendung von subjektiven Einschätzungen zur Erhebung von Workload entstehen können. So ist es möglich, dass die Operateure physische Belastungen und mentale Beanspruchung in der Bewertung konfundieren könnten. Außerdem können Nutzer nur schwer zwischen der Aufgaben-Schwierigkeit und der eigentlichen Workload unterscheiden. Weiterhin ist davon auszugehen, dass nicht alle Aspekte der empfundenen Belastung und Beanspruchung anhand der Introspektion durch den Nutzer zugänglich sind und bemessen werden können. Dies hat auch zur Folge, dass Workload-Messungen zeitnah zu erheben sind, um keinen Informationsverlust auf Grund des Kurzzeitgedächtnisses zu provozieren. Als weiteren Punkt führen die Autoren die Dissoziation subjektiver Maße und der Messung von Nutzerleistung während der Durchführung der Primär-Aufgabe an.

Zu einer Dissoziation von subjektiven Maßen und der Nutzerleistung kommt es, wenn die subjektive Bewertung der Workload durch den Nutzer nicht mit den Maßen der Nutzerleistung konform geht (Yeh & Wickens, 1984). An dieser Stelle wird deutlich, dass Nutzerleistung und Workload durch unterschiedliche Faktoren beeinflusst werden und auch die Messmethoden von Faktoren beeinträchtigt werden können, die nicht direkt mit Workload in Zusammenhang stehen (Hart & Wickens, 1990). Diese Implikationen sind daher bei der Verwendung von subjektiven Maßen und der Erhebung von Nutzerleistung zu beachten und sollten bei der Interpretation der Ergebnisse berücksichtigt werden.

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

Im Zweitaufgaben-Paradigma (s.a. Eichinger, 2011; Kellerer, 2010) kann die Nutzerleistung anhand von objektiven Werten während der Durchführung der zu untersuchenden Haupt- oder Primäraufgabe zum Beispiel durch die Performance Operating Characteristic (Eichinger, 2011) gemessen werden; hierbei wird angenommen, dass der Nutzer während der Durchführung der Aufgabe auf Grund einer sich verändernden Aufgaben-Schwierigkeit seine Handlungsstrategien anpasst und er mehr Anstrengung investieren muss, um diese Aufgabe weiterhin erfolgreich zu erfüllen (Wierwille et al., 1979). Eine weitere Herangehensweise ist es, die Bearbeitung der Aufgabe selbst als Indiz zur Bestimmung von Workload zu verwenden und Einbußen in der Leistung oder ein erfolgloses Abschließen der Aufgabe als Hinweis für eine zu hohe Nutzer-Workload anzunehmen (Wierwille et al., 1979). Lysaght, Hill, Dick, Plamondon und Linton (1989) unterscheiden zwei Formen der Erhebung zur Bestimmung der Belastung und Beanspruchung des Nutzers: Messungen zur Beschreibung des Gesamt-Systems unter Betrachtung der zu erfüllenden Aufgabe und direkte Maße der Nutzerleistung des Operators. Beide Vorgehensweisen sollten hierbei zur gegenseitigen Erklärung genutzt werden, da die Nutzerleistung nur im Kontext der Leistung des Gesamt-Systems erklärt werden kann. Das heißt, das auch die Nutzerleistung nicht als Einzelmaß verwendet werden sollte (Cain, 2007).

Als mögliche Maße in der Erhebung von Leistungen in den Primär-Aufgaben können nach Hart und Wickens (1990) beispielsweise die Anzahl korrekter Antworten während der Durchführung der Nutzeraufgabe, die Anzahl von Fehlern des Nutzers, Reaktionszeiten oder aber auch der Erfüllungsgrad der Aufgabe verwendet werden. Die Erhebung von objektiver Nutzerleistung ist einfach umzusetzen und wird in der Regel in den meisten Studien miterhoben (Stanton et al., 2011). Speziell in Situationen, in denen die Workload des Nutzers überhandnimmt und deshalb mit Einbußen in der Hauptaufgabe des Nutzers zu rechnen ist, sind diese Maße geeignet (Cain, 2007). Gopher und Donchin (1986) bewerten die Erhebung von Nutzerleistung als Maß für Workload als ungeeignet, da diese meist nicht kennzeichnen, wie sehr der Ressourcen-Einsatz der Nutzer variiert, auf welche Form der Modalität die Belastung und Beanspruchung des Nutzers zurückzuführen ist und weiterhin kann kein zuverlässiger Übertrag auf die Leistung des Gesamt-Systems stattfinden. Hart und Wickens (1990) empfehlen daher die Leistungsmessungen der Primär-Aufgabe lediglich dazu zu nutzen, um festzustellen, ob die Aufgabe erfüllt wurde, da mit diesen Messungen Workload unter Umständen nicht oder nur indirekt erhoben wird. Eichinger (2011) hingegen zeigt unter Verwendung der Performance Operating Characteristic, bei der zwei Aufgaben gegenübergestellt werden, dass die Leistungseinbußen jeweils in Abhängigkeit einer zweiten Aufgabe graphisch dargestellt werden können.

Die Erhebung von Workload im Rahmen des Zweitaufgaben-Paradigmas erhebt die Nutzerleistung der Probanden im Sinne freier mentaler Kapazität und erlaubt so einen Einblick hinsichtlich der vom Nutzer benötigten Ressourcen (Hart & Wickens, 1990). Eine Anwendung dieses Vorgehens geschieht, indem neben der Hauptaufgabe – also der Durchführung der Flugaufgabe, während der Pilot die Cockpit-Modifikation verwendet, eine zusätzliche Aufgabe durch den Operator zu bearbeiten ist und hierin dessen Leistung erhoben wird. Ein Abfallen der Nutzerleistung in der Zweitaufgabe wird schließlich im Sinne von Einbußen der freien mentalen Kapazität gewertet (Kantowitz & Casper, 1988). Dies ist deshalb zielführend, weil angenommen wird, dass der Pilot nur die Ressourcen auf die

Zweitaufgabe verwendet, die dieser neben der Bearbeitung der Hauptaufgabe erübrigen kann und somit auch die Belastung und Beanspruchung des Nutzers durch die Hauptaufgabe in der Bearbeitung der Nebenaufgabe widerspiegelt wird (Vidulich, 2003). Bei der Durchführung der Zweitaufgabe ist daher der Operateur dahingehend zu instruieren, dass die Hauptaufgabe vom Nutzer bestmöglich erfüllt wird und die Zweitaufgabe mit den verbleibenden Kapazitäten zu bearbeiten ist. Als typische Zweitaufgaben gelten Gedächtnisabruf, Rechenaufgaben, Reaktionszeiten und Nachverfolgungs-Aufgaben (Stanton et al., 2011), aber auch Signalentdeckungs-Aufgaben (Cain, 2007). O'Donnell und Eggemeier (1986) führen weiterhin auch Überwachungsaufgaben und Aufgaben zur Zeiteinschätzung an.

Das Zweitaufgaben-Paradigma eignet sich, um die verfügbare Kapazität der Nutzer zu erheben und auf dieser Grundlage verschiedene Aufgaben oder eben auch verschiedene Konfigurationen der zu untersuchenden Cockpit-Modifikation zu beschreiben, sofern diese durch Nutzerleistung unterscheidbar sind (Gawron, 2008). Bei der Erhebung von Workload anhand des Zweitaufgaben-Paradigmas ist allerdings problematisch, dass nicht kontrolliert werden kann, wieviel Aufmerksamkeit der Operateur der Zweitaufgabe widmet und weiterhin das Ausmaß der Intrusivität der Zweitaufgabe, was wiederum die Leistungen in der Hauptaufgabe beeinflussen kann (Wickens et al., 2013). Hinzu kommt, dass die Nutzer Strategien entwickeln können, um die Belastungen der Zweitaufgabe zu mitigieren, indem sie nach Schneider und Detweiler (1988) beispielsweise die Bearbeitungsstrategie verändern, die Aufgaben versetzt bearbeiten oder sich auf die Bearbeitung der Aufgaben vorbereiten.

Ein weiteres Messvorgehen zur Bestimmung von Nutzer-Workload ist die Verwendung von psycho-physiologischen Maßen. Hier werden die Indikatoren für mentale Beanspruchung registriert, die mit der Aufgabenkomplexität und deren Bewältigung kovariieren (Manzey, 1998). Angenommen wird, dass bei Änderungen der Workload auch psycho-physiologische Auswirkungen auf den menschlichen Körper messbar werden und dass diese Variablen erhoben werden können (Wierwille et al., 1979). Dies können hirnelektrische Indikatoren, kardiovaskuläre Indikatoren, Indikatoren der Atmung, Indikatoren des okulomotorischen Systems, aber auch andere Indikatoren wie tonische Muskelaktivität, elektrodermale Aktivität oder Reaktionen des endokrinen Systems sein (Manzey, 1998). Für die Erhebung von Nutzer-Workload in der Luftfahrt bezeichnet Wierwille (1979) die Erhebung von Pupillen-Dilatation, die Erhebung von ereigniskorrelierten Potenzialen und die Messung von Körperflüssigkeiten und deren Bestandteile in Speichel oder Urin als die vielversprechendsten Maße. Die Erhebung von psycho-physiologischen Messungen ist insofern vorteilhaft, als dass sie eine kontinuierliche Messform mit objektiven Maßen darstellt, welche die eigentliche Bearbeitung der Nutzeraufgabe in der Regel kaum stört (Hart & Wickens, 1990). Die meisten psycho-physiologischen Messungen scheinen aber weniger konsistent mit Workload zu kovariieren, was an der Verrauschtheit dieser Signale liegen kann, wie es beispielsweise anhand des Einflusses von Beleuchtung auf die Pupillen-Dilatation offenkundig wird (Kramer, 1990) und zu einer geringeren Reliabilität führt. Aus diesem Grunde sollten sie nicht als einzige Maße zur Erhebung von Workload verwendet werden (Hart & Wickens, 1990). Psycho-physiologische Messungen können aber beispielsweise zur Ergänzung von

Nutzerleistungen und subjektiven Erhebungen der Nutzer-Workload hinzugezogen werden, wenn diese uneindeutig sind oder eine Veränderung der Nutzerstrategien aufgedeckt werden soll. Psychophysiologische Messungen der Workload geben für sich genommen aber eher darüber Auskunft, wie der Körper des Nutzers auf die Auswirkungen der empfundenen Workload reagiert (Cain, 2007).

In der Zusammenfassung und Abwägung der beschriebenen Strategien zur Erhebung von Nutzer-Workload stellt sich heraus, dass kein Vorgehen eindeutig bevorzugt werden kann. Auch Williges und Wierwille (1979) kommen nach Abwägung von 14 Vorgehen zur Erhebung von Nutzer-Workload im Bereich der Luftfahrt mittels Verhaltensmaße zu dem Schluss, dass Workload auf Grund seiner Multidimensionalität nur anhand verschiedener Messstrategien erhoben werden sollte.

3.4.2.1 NASA Task Load Index (TLX)

Der NASA TLX ist ein subjektiver, multidimensionaler Fragebogen zur Erfassung von Workload (Hart & Staveland, 1988; Hart, 2006), der ursprünglich für die Luftfahrt entwickelt wurde und auch in anderen Branchen auf breite Akzeptanz stößt (Hart, 2006). Workload wird mit dem NASA TLX anhand von sechs Dimensionen bzw. Subskalen operationalisiert, die sich in zwei Kategorien untergliedern lassen – Anforderungsskalen und Verhaltensskalen (Eichinger, 2011). Unter den Anforderungsskalen versteht man hier die Skalen kognitive Anforderungen, physische Anforderungen und zeitliche Anforderungen, während die verhaltensbasierten Skalen die Dimensionen Aufwand, Leistung und das vom Nutzer empfundene Level an Frustration umfassen. Jeder dieser Aspekte wird anhand von einer spezifischen Subskala bestimmt, die nach Hart und Staveland (1988) eine eher große Spannweite von 10 bis 20 einnehmen soll und mit Ausnahme der Subskala ‚Leistung‘ mit den Endpunkten niedrig und hoch bewertet werden soll. Die Subskala ‚Leistung‘ hingegen wird mit den Ankern ‚gut‘ und ‚schlecht‘ erhoben. Rabl (2016) führt an, dass auch eine Normierung in einen anderen Zahlenraum wie beispielsweise zwischen 0-10 oder 0-100 ohne weiteres möglich ist.

Bei Verwendung dieser Methode ist nach Abschluss der Workload-Einschätzung auch eine Gewichtung der verschiedenen Dimensionen durch den Nutzer vorgesehen, um der Bedeutung situativer und individueller Aspekte gerecht zu werden und auf dieser Grundlage einen gewichteten Workload-Gesamtwert zu errechnen (Hart & Staveland, 1988; Hart, 2006). Hierzu wird der Nutzer nach seiner Workload-Einschätzung anhand von 15 Paarvergleichen dazu befragt, wie stark ihn die unterschiedlichen Dimensionen während der Durchführung seiner Aufgabe beeinträchtigt haben. Das heißt, dass der befragte Nutzer entscheiden muss, ob er beispielsweise kognitiv oder zeitlich stärker gebunden war. Diese Form des Paarvergleichs ist dann für jede weitere mögliche Kombination der Workload-Dimensionen durchzuführen. Die Gewichtung entsteht, indem gezählt wird, wie oft eine Subskala als belastender eingestuft wird als eine andere. Für eine Workload-Dimension ergibt sich somit eine maximal mögliche Gewichtung von fünf. Durch Einführung der Gewichtung wird berücksichtigt, dass Workload von den befragten Nutzern auch unterschiedlich empfunden wird. Die Berechnung des Gesamt-Workloades erfolgt schließlich, indem der Skalenwert mit der Gewichtung multipliziert, alle multiplizierten Skalenwerte addiert und dieser Wert durch das Gesamtgewicht von 15 dividiert wird (Hart & Staveland, 1988). In einem Übersichtsartikel stellt Hart (2006) fest, dass es

eine gängige Anpassung des NASA TLX ist, die Gewichtung wegzulassen und den NASA TLX als Raw TLX zu verwenden. Ein Vergleich der gewichteten und der ungewichteten Gesamtwerte führt hierbei zu unterschiedlichen Ergebnissen und ein Weglassen der Gewichtung scheint insofern legitim. Schon Byers, Bittner und Hill (1989) stellen im Rahmen von sechs Studien fest, dass hohe Korrelationen zwischen den TLX-Werten sowie deren Mittelwerten und Standardabweichungen bei einer Verwendung des Fragebogens mit und ohne Gewichtung bestehen und empfehlen daher die Verwendung des Raw TLX. Nygren (1991) hingegen rät von der Verwendung des Gewichtungsverfahrens regelrecht ab, da dieses ineffektiv ist und zu suboptimalen psychometrischen Eigenschaften des NASA TLX führen kann, weil die Gewichtung nur zu einer relativen Ordnung der Dimensionen führt, in der es zu keiner echten Gewichtung der Skalen kommt. Außerdem könne eine Skala den Wert 0 annehmen und andere lineare Gewichtungen können zu einer besseren Schätzung gelangen, als dies durch die Paarvergleiche der Fall ist.

Für das multidimensionale Methoden-Inventar ist der NASA TLX auf Grund seiner Bekanntheit und seiner schnellen Anwendbarkeit geeignet, insbesondere aber, weil der NASA TLX hauptsächlich in der Luftfahrt validiert wurde. Die Anwendung des NASA TLX zur Beschreibung von Workload kann hierbei während oder nach Durchführung einer Nutzeraufgabe erfolgen (Hart, 2006). Für eine Anwendung dieses Fragebogens als Online-Methodik – also der Applikation, während der Pilot eine Nutzeraufgabe zur Überprüfung einer technischen Funktion durchführt, ist diese Methode jedoch als zu intrusiv anzusehen.

3.4.2.2 *Tactile Detection Response Task - TDRT*

Eine spezifische Form der Zweitaufgabe ist die Detection Response Task (DRT), die sich als zuverlässige Methode zur Erhebung von dynamischer Workload in Fahr simulatoren bewährt hat und daher auch als normierte Aufgabe zur Erfassung von Workload im Fahrzeugkontext genutzt wird. Nach der Norm ISO 17488:2016 ist unter einer DRT eine Aufgabe zu verstehen, bei der zur Erhebung der freien Kapazitäten des Operators, dessen Reaktionszeiten und Trefferquoten auf einen artifiziellen Reiz erfasst werden. Da es sich um eine Aufgabe im Sinne des Zweitaufgaben-Paradigmas handelt, ist allerdings auch eine Auswertung als ‚Performance Operating Characteristic‘, wie sie Eichinger (2011) vorschlägt, möglich, in welcher die Nutzerleistungen zweier gleichzeitig durchgeführter Aufgaben gegenübergestellt werden.

Die Reizapplikation kann entweder visuell, auditiv oder taktil erfolgen und soll relativ häufig in einem festgelegten Zeitintervall mit einem zu bestimmenden Unsicherheitsgrad (Reizapplikation alle drei bis fünf Sekunden) stattfinden. Das Antwortverhalten durch den Nutzer wird per einfachen Knopfdruck (Micro-Switch) durch den Nutzer aufgezeichnet und soll dem Nutzer nach Eingabe eine erkennbare Rückmeldung zur Verfügung stellen. Aufgrund der hohen visuellen Belastungen ist die Anwendung einer taktilen Form der DRT (TDRT) im Cockpit ratsam, da so Interferenzen mit der visuellen Workload des Nutzers vermieden werden können (ISO 17488:2016).

Studien, die der Normierung der DRT vorausgegangen sind, untermauern die Diagnostizität und Sensitivität der DRT im Allgemeinen und auch der TDRT im Spezifischen. So konnten Young, Hsieh

und Seaman (2013) beispielsweise in einer Laborstudie sowie in der praktischen Anwendung herausfinden, dass die TDRT auf einen Anstieg der mentalen Workload reagiert, nicht jedoch bei einem spezifischen Anstieg einer rein visuellen Beanspruchung. Harbluk, Burns, Hernandez, Tam und Glazduri (2013) ergänzen, dass die TDRT im Vergleich der angewendeten DRTs als einzige Form dazu geeignet war, neben der induzierten Workload auch zur Baseline hin zu unterscheiden und weiterhin in ihrer Anwendung leicht am Nutzer applizierbar ist und nicht mit etwaigen Eye-Tracking-Messungen interferiert. Da die DRT eine genormte Methode zur Erfassung von Aufmerksamkeitsbindung im Straßenverkehr ist, ist anzunehmen, dass diese auch sensitiv auf Workload-Veränderungen im Cockpit-Simulator eines hochagilen Luftfahrzeuges reagiert und aufgrund ihrer taktilen Anwendbarkeit für den Cockpit-Simulator geeignet erscheint.

3.4.2.3 Eye-Tracking

Eine Form des psycho-physiologischen Messens der Workload ist mit der Verwendung des Eye-Trackings möglich. Die Erfassung der Workload kann mittels okulomotorischer Messungen wie Lid-schlüsse, Pupillenreaktionen sowie auch durch Maße des Blickverhaltens vorgenommen werden (Grandt, 2004).

Bezüglich des Blickverhaltens besteht im Luftfahrtkontext die Annahme, dass die Länge der Verweildauer einer Fixation einen Einfluss auf die visuelle Workload der Piloten hat, da davon auszugehen ist, dass beispielsweise Displayinhalte schwieriger zu lesen sind, wenn der Nutzer länger darauf blickt (Lysaght et al., 1989). In einer Untersuchung zum visuellen Scan-Verhalten von Piloten unter Belastung in einer simulierten Nutzeraufgabe stellen Harris, Tole, Stephens und Ephrath (1981) fest, dass mit zunehmender Flugfertigkeit die Tendenz zu überlangem Verweilen auf Displayinhalte abnimmt. Auch Moore und Gugerty (2010) zeigen, dass mit einem Anstieg der Verweildauer auf einige wenige Displayinhalte weniger Zeit verbleibt, um verbleibende Tätigkeiten durchzuführen. Diese Form der Aufmerksamkeitsfixierung könnte daher ein Indiz für den Anstieg von Workload sein, wie es O'Donnell und Eggemeier (1986) beschreiben, da der Nutzer seine Informationssuche auf Grund des zeitlichen Drucks anpassen muss.

Eine weitere Messgröße zur Beschreibung der Workload ist das Maß der Pupillen-Dilatation, das bereits von Kahneman (1973) als geeignete Messgröße für Workload angesehen wird, da diese sensitiv auf Unterschiede innerhalb einer (Nutzer-) Aufgabe reagiert, wie auch zwischen zwei verschiedenen Aufgaben unterscheiden kann. Nach Manzey (1998) korreliert die Amplitudenhöhe der phasischen Pupillenreaktion mit der Schwierigkeit oder Komplexität der Aufgabe und zeigt daher Workload an. Die Schwierigkeit bei der Erhebung der Pupillengröße liegt allerdings darin, dass die Pupillen-Dilatation durch andere Effekte wie Lichteinstrahlung zum einen und der Fokuswechsel von entfernten Objekten zu nahen Objekten zum anderen stärker beeinflusst wird als dies durch die Beeinflussung mentaler Workload der Fall ist (Kramer, 1990). Das von Kramer (1990) angesprochene Problem kann insbesondere für die Anwendung in der Simulationsumgebung eine Herausforderung sein, da es hier notwendig ist, für konstante Lichtverhältnisse zu sorgen. Eine mögliche Lösung für dieses Problem ist die Verwendung des Index of Cognitive Activity (ICA), der von Marshall (1999) entwickelt

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

wurde und lediglich schnelle und spontane Pupillenerweiterungen mit Hilfe einer Wavelet-Analyse herausfiltert und daher weniger empfindlich auf Lichteinflüsse reagieren soll. Demberg (2013) beschreibt allerdings in einer Studie zu einer simulierten Fahraufgabe mit eingesetzter Sekundär-Aufgabe den ICA von denen Messungen der Pupillen-Dilatation als verschieden, da diese kaum miteinander korrelieren. Beide Messungen reagieren jedoch sensitiv auf die Zweitaufgaben, wenn auch nicht auf die Primär-Aufgabe.

Neben der Pupillen-Dilatation wird auch die Abnahme der Frequenz von spontanen Lidschlüssen als Indiz für Workload bei visuell zu bearbeitenden Aufgaben gesehen, da dieses Maß sensitiv auf eine Veränderung der Aufgaben-Schwierigkeit bei visuellen, jedoch nicht bei auditiven oder arithmetischen Zweitaufgaben reagiert (Manzey, 1998). Eine weitere Schwierigkeit bei der Verwendung von Lidschlüssen zur Erhebung von Workload ist, dass diese sowohl auf hohe visuelle Workload wie auch auf die Ermüdung der Operateure reagieren (Caldwell et al., 1994).

Eine Möglichkeit zur Erhebung von Müdigkeit ist daher folgerichtig das Maß PERCLOS, das Wierwille, Ellsworth, Wreggit, Kirn und Fairbanks (1994) im Hinblick auf andere Ermüdungs- und Leistungsmaße im Fahrzeugkontext validieren. Der PERCLOS als Müdigkeitsmaß beschreibt hierbei den zeitlichen Anteil der Nutzeraufgabe, in welchem das Augenlid zu mehr als 80 % geschlossen ist. Allerdings stellen Bowman und Hanowski (2008) fest, obwohl sie den PERCLOS als ‚Goldstandard‘ zur Erhebung von Ermüdung zitieren, dass eine Kombination des PERCLOS mit anderen, leistungsbasierten Messungen im Fahrzeugkontext zu stabileren Ergebnissen in der praktischen Anwendung führen kann, da so die technischen Schwierigkeiten dieser Methodik (beispielsweise Datenverlust bei Brillenträgern) wie auch der leistungsbasierten Messungen ausgeglichen werden können. Sommer und Golz (2010) führen allerdings auf Grund der Ergebnisse ihres Labor-Experiments aus, dass für eine Unterscheidung von geringer und starker Müdigkeit eine Kombination aus Elektro-Okulogramm und Elektro-Enzephalogramm Messungen eindeutigere Ergebnisse erzielt als der PERCLOS.

3.4.3 Methodenauswahl zur Erfassung von Gebrauchstauglichkeit

Zur Überprüfung der Gebrauchstauglichkeit der Cockpit-Modifikationen durch Implementierung neuer technischer Funktionen ist es nötig, die Mensch-Maschine-Schnittstelle zielführend und mittels Nutzereinbindung zu testen, da hier der Nutzer direkt betroffen ist (J. Turner, 1997). Diese Überprüfung geschieht, da es sich in erster Linie um Bedienoberflächen und Anzeigen handelt, insbesondere hinsichtlich der Kriterien Effizienz, Effektivität und Nutzerzufriedenheit, die in der Norm DIN EN ISO 9241-210 festgelegt sind. Das objektive Maß Effektivität wird demnach als Genauigkeit und Vollständigkeit verstanden, mit der Nutzer ihr Nutzungsziel erreichen, während Effizienz den Ressourceneinsatz kennzeichnet, um dieses Maß an Aufgabenerfüllung zu erreichen. Zufriedenstellung als subjektives Maß hingegen beschreibt die positive Einstellung des Nutzers gegenüber dem Produkt sowie die Freiheit des Produktes von Beeinträchtigungen, die sich auf den Nutzer auswirken (DIN EN ISO 9241-210). Quesenbery (2004) verwendet die Aspekte Effektivität und Effizienz und erweitert die Definition der Gebrauchstauglichkeit anstatt der Verwendung des Begriffs ‚Nutzerzufriedenheit‘ um

drei weitere Dimensionen: Ansprechbarkeit des Designs (s.a. Norman, 2002), Fehlertoleranz und Erlernbarkeit des Systems.

Zur praktischen Erhebung der Gebrauchstauglichkeitskriterien führen Shneiderman et al. (2005) folgende zielführende Überlegungen an, deren Bedeutung jeweils an den Nutzungskontext anzupassen ist:

- *Erlernbarkeit*: Wie lange benötigt der Nutzer, die Handlungen des Systems zu erlernen?
- *Bedienzeit*: Wie lange benötigt der Nutzer, um die erforderlichen Aufgaben zu erfüllen?
- *Fehlerrate*: Wie viele Fehler treten bei der Bedienung des Systems auf und welcher Art sind diese? Können Fehler korrigiert werden (z.B. durch eine UNDO-Funktion)?
- *Eingängigkeit*: Wie robust ist die Erinnerung der erlernten Bedienschritte?
- *Subjektive Nutzerzufriedenheit*: Wie zufriedenstellend ist die Bedienung des Systems für den Nutzer?

Um die genannten Fragen zu beantworten, werden nach Sauro und Lewis (2009) von den meisten Anwendern Maße erhoben, die den Erfüllungsgrad der Aufgabe beschreiben, die aufgewendete Bedienzeit der Nutzer, auftretende Fehlerraten sowie die Zufriedenheit der Nutzer mit dem jeweiligen Erfüllungsgrad der Aufgabe und des Tests. Sauro und Lewis (2012) führen als weitere Maße die Häufigkeit von Hilfsmaßnahmen und -mitteln sowie ein Auflisten an Nutzerproblemen an. Um diese Kriterien zu erheben, kann ein formatives oder ein summatives Vorgehen gewählt werden, wobei sich eine formatives Vorgehen auf einen Systemaspekt bezieht, der noch in der Entwicklung ist und dahingehend Fehler im Design entdeckt werden sollen, während sich ein summatives Testen auf ein fertiges Produkt bezieht, welches validiert und hinsichtlich der Anforderungserfüllung überprüft werden soll (s.a. Barnum, 2011). Die Überprüfung der Gebrauchstauglichkeit der Cockpit-Modifikation ist daher im Bereich des formativen Vorgehens anzusiedeln, da weitere iterative Entwicklungsschritte bis zur vollständigen Implementierung in das System notwendig sind. Bei der Abwägung der Analysemethoden zur Beschreibung der Gebrauchstauglichkeit der MMS stellt sich heraus, dass diese in der Regel leicht verständlich, schnell anwendbar und kostengünstig sind, jedoch zeit- und arbeitsintensive Auswerteverfahren notwendig machen (Stanton et al., 2011).

3.4.3.1 Nutzerleistung

Unter Nutzerleistung versteht man nach dem Technical Report ISO/ TR16982:2002(E) quantifizierbare und aufgabenbezogene Leistungsmaße des Nutzers, die dazu genutzt werden, um Effektivität und Effizienz zu erheben. Hierunter fallen beispielsweise die Bedienzeit bis zum erfolgreichen Abschluss der Aufgabe, die Anzahl der durchgeführten Aufgaben in einem Zeitintervall, die Anzahl an Bedienfehlern oder auch die Zeit, die der Nutzer benötigt, um eine bestimmte Information zu erfassen und zu interpretieren.

Nach Albert und Tullis (2013) sind Effektivität und Effizienz für die Erhebung der Gebrauchstauglichkeit und deren Bestimmung anhand der Nutzerleistung deshalb so wichtig, da so beschrieben werden kann, wie genau der Nutzer die Aufgabe bearbeitet (beispielsweise anhand der Fehlerquote)

und welchen Ressourcen-Einsatz (beispielsweise Zeit) er hierfür benötigt. Es kann also das Ausmaß eines Problems der Gebrauchstauglichkeit erfasst werden. Die Autoren warnen aber davor, sich allein auf die Leistungsdaten der Nutzer zu verlassen und empfehlen auch, diese Maße durch andere Messvorgehen zu ergänzen.

Der Vorteil der Erfassung von Nutzerleistung ist die Erhebung von objektivem Datenmaterial, das leicht gewonnen, interpretiert und anschließend mit anderen Gestaltungsvarianten verglichen werden kann. Nachteilig hingegen ist, dass die erhobenen Daten nicht zwangsläufig die Ursache für die zu Grunde liegenden Probleme der Gebrauchstauglichkeit offenlegen und für die Erhebung der Maße Prototypen vorhanden sein müssen (Technical Report ISO/ TR16982:2002(E)).

3.4.3.2 *Post Study System Usability Questionnaire – PSSUQ*

Der PSSUQ ist ein standardisierter Fragebogen, der ursprünglich für die Erhebung der empfundenen Nutzerzufriedenheit bezüglich Computersysteme und Applikationen entwickelt wurde. Er basierte ursprünglich auf 18 Items, die sich nach Durchführung eines Projekts und einer hierauf basierenden Inhaltsanalyse von offenen bzw. teilstandardisierten Interviews herauskristallisierten (Sauro & Lewis, 2012). Die verwendeten Items umfassen hier klassische Gebrauchstauglichkeits-Maße wie die Leichtigkeit der Systemnutzung, die Leichtigkeit des Erlernens und Einfachheit des Systems sowie Fragen zu Informationsqualität der Mensch-Maschine-Schnittstelle (J. R. Lewis, 1992). Der Fragebogen wurde erstmals 1992 von J. R. Lewis validiert. Er kommt nach einem Vergleich mit einem weiteren Fragebogen zu dem Schluss, dass diese das gleiche Konstrukt Gebrauchstauglichkeit messen. Nach weiteren Versionen und Validierungsschritten wird aktuell die Nutzerzufriedenheit mittels PSSUQ anhand von 16 positiv formulierten Aussagen erhoben, welche den Faktoren ‚Nützlichkeit des Systems‘, ‚Informationsqualität‘ und ‚Qualität der Schnittstelle‘ sowie insgesamt der zusammenfassenden Gesamtgebrauchstauglichkeit zugeordnet werden können. Die Aussagen selbst werden anhand einer siebenstufigen Rating-Skala mit den Ankern 1 (stimme voll zu) und 7 (stimme überhaupt nicht zu) erfasst, wobei niedrigere Werte einen höheren Grad an Zufriedenheit ausdrücken (Sauro & Lewis, 2012). Bei einem Vergleich der ursprünglichen Fassung des PSSUQ und der derzeit angewendeten Form stellte sich heraus, dass die Faktorenstruktur, Skalenreliabilität und Sensitivitätsanalyse zu vergleichbaren Ergebnissen führt, wobei die ‚Qualität der Schnittstelle‘ als Faktor am meisten Varianz aufweist (J. R. Lewis, 2002). Auch Fruhling und Lee (2005) konnten in ihrer Studie die Reliabilität des PSSUQ aufzeigen, berichten aber nicht von drei relevanten Faktoren, sondern von zwei Faktoren (‚Nützlichkeit des Systems‘ und ‚Qualität des Systems‘) bei Anwendung der ursprünglichen PSSUQ-Version. Weiterhin zeigt sich der PSSUQ als insensitiv hinsichtlich der Beantwortung des Fragebogens durch verschiedene Geschlechter und gilt außerdem als robust, selbst wenn nicht alle Items geratet werden. Allerdings raten die Autoren davon ab, den Fragebogen kulturübergreifend zu verwenden, da hier Unterschiede in den Antwort-Tendenzen entstehen könnten (J. R. Lewis, 2002).

3.4.3.3 *Eye Tracking*

Die Verwendung von Eye-Tracking als Messmethodik im Cockpit ist deshalb besonders sinnvoll, da die visuelle Präsentation von Informationen erfasst werden kann, ohne hierbei intrusiv zu sein (Peißl, Wickens & Baruah, 2018).

Als Messmethodik zur Bestimmung von Gebrauchstauglichkeit kann Eye-Tracking dazu genutzt werden, um beispielsweise die Bedienbarkeit und Verteilung von Informationen verschiedener Schnittstellen zu untersuchen. Goldberg und Kotval (1998) nutzen hierzu verschiedene Bewegungsparameter des Auges und Scan-Pfade zur Unterscheidung von prototypischen ‚guten‘ und ‚schlechten‘ Mensch-Maschinen-Schnittstellen am Beispiel eines Zeichenprogramms. Es zeigte sich, dass ein Design, in welchem die Informationszuordnung nicht strukturiert war, zu einer höheren Anzahl an Sakkaden bei gleichbleibender Länge führte. Als weitere Messung zur Beschreibung der Schnittstellenqualität kann die Sakkaden-Amplitude gewählt werden. Allerdings zeigte sich in der Untersuchung kein Unterschied. Zwar war das Suchverhalten der Probanden in dem Design mit schlechter Qualität intensiver, die gemittelte Amplitude zeigte jedoch keine Unterschiede. Die Autoren folgern weiterhin aus ihrer Untersuchung, dass Eye-Tracking und die Untersuchung von Bewegungsparametern Nutzerstrategien aufzeigen und somit als sinnvolle Ergänzung zur Erhebung von Gebrauchstauglichkeit genutzt werden können, diese Methodik jedoch nur einen Teil der hierunter zusammengefassten Themen erfassen kann.

Ein weiterer Indikator zur Bestimmung der Gebrauchstauglichkeit könnten Zielfindungsmessungen sein, da diese die Effektivität und Effizienz einer Suchstrategie erfassen, indem sie sowohl den Aspekt des Suchens als auch den des Identifizierens durch den Nutzer festhalten. Allerdings hat der Einsatz derlei Messungen nur dann Sinn, wenn ein bestimmtes Objekt der Nutzeroberfläche erfasst werden soll (Bojko, 2013).

Zur Bestimmung von Gebrauchstauglichkeit von Cockpit-Modifikationen in der Simulationsumgebung ist das Eye-Tracking als Messmethodik interessant, da dieses die Nutzerstrategie und dessen Suchverhalten verdeutlicht, ohne hierbei intrusiv zu sein. Goldberg und Wichansky (2003) sowie Holmqvist et al. (2011) kritisieren allerdings, dass für die Verwendung von Eye-Tracking-Maßen zur Bestimmung der Gebrauchstauglichkeit keine Standards gegeben sind und Untersuchungen durchgeführt werden sollten, die einen Zusammenhang von Eye-Tracking-Maßen und den standardisierten Gebrauchstauglichkeits-Maßen (Effektivität, Effizienz und Nutzerzufriedenheit) herstellen.

3.4.3.4 *Interviews / Experten-Befragungen*

Interviews als Methode sind anzuwenden, wenn Informationen, Fakten und Meinungen erfragt werden sollen (Allhoff & Allhoff, 2010). Zur Analyse von Cockpit-Modifikationen können Interviews als eine flexible und einfache Form der Daten-Erhebung genutzt werden, um spezifische Aspekte hinsichtlich der Nutzung der Cockpit-Modifikationen zu erfragen, die nicht mit Fragebögen zur Nutzerzufriedenheit und Nutzerleistung abgedeckt werden können. Hier können spezifische Design-Entscheidungen diskutiert und die subjektiven Meinungen der Experten besprochen werden. Durch

die Anwendung von offenen und teilstrukturierten Interviews ist es also das Ziel, subjektive Erklärungen sowie komplexe kognitive Strukturen und deren Veränderlichkeit aufzudecken (Wottawa & Thierau, 1998). Interviews werden nach Bortz und Döring (2006) hinsichtlich dem Ausmaß an Standardisierung, nach Autoritätsanspruch, Zugang zum Interviewpartner, Anzahl von Interviewer und Interviewten und hinsichtlich ihrer Funktion unterschieden. Die Zuordnung der Interviews zur Analyse von Cockpit-Modifikationen ist hier als direktes, semi-strukturiertes, informatorisches Einzelinterview anzusehen, da der Experte vor Ort anhand von vorstrukturierten, offenen Fragestellungen befragt werden soll, um dessen Fachwissen und subjektive Meinung zu erfassen. Die verwendete Interview-Form ist auch deshalb teil-strukturiert, weil es dem interviewten Experten möglich sein soll, nach eigenem Ermessen und Einschätzung zu antworten und dem Befragenden die Möglichkeit eingeräumt werden soll, klärende Nachfragen zu stellen (Hopf, 1995). Die angewendete Form des Interviews orientiert sich des Weiteren an der spezifischeren Definition des systematisierenden Experteninterviews, da es sich um die Erhebung von exklusivem Expertenwissen in Form von Handlungs- und Erfahrungswissen handelt, welches mit Hilfe eines ausdifferenzierten Leitfadens erhoben werden soll und ein zentrales Erhebungsinstrument in multimethodischen Ansätzen ist (Bogner & Menz, 2002).

3.4.3.5 Evaluation Checklist

Ravden und Johnson (1989) untersuchten unterschiedliche Schnittstellen in der Fertigungsindustrie jeweils zu verschiedenen Entwicklungszeitpunkten und entwickelten iterativ und unter Einbindung der Richtlinien zur Schnittstellengestaltung die Evaluation Checklist (Johnson, 1996). Anhand dieser Methodik sollen repräsentative Aufgaben durch den Nutzer getätigt und anschließend mittels der zu bewertenden Aspekte der Checkliste von Nutzern mit Systemerfahrung evaluiert werden. Die Anwendung der Evaluation Checklist verfolgt das Ziel, Problembereiche und Stärken der zu untersuchenden Schnittstelle aufzuzeigen, um darauf aufbauend die Gebrauchstauglichkeit zu verbessern. Die Evaluation Checklist nach Ravden und Johnson (1989) besteht aus neun Zielen des Schnittstellendesigns, die jeweils anhand detaillierter Fragestellungen spezifiziert werden sollen.

- *Visual clarity*
Der Displayinhalt soll leicht lesbar, eindeutig und klar strukturiert sein.
- *Consistency*
Das System soll in Verhalten und Aussehen konsistent sein.
- *Compatibility*
Das Verhalten des Systems soll den Erwartungen des Nutzers entsprechen.
- *Informative feedback*
Der Nutzer soll eindeutige Rückmeldungen zu Status, Position und Handlungen erhalten.
- *Explicitness*
Arbeitsweise und Struktur des Systems müssen für den Nutzer klar erkennbar sein.
- *Appropriate functionality*
Das System muss den Anforderungen des Nutzers genügen, damit dieser seine Aufgaben erfüllen kann.

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

- *Flexibility and control*
Das System soll dem Nutzer die benötigte Kontrolle ermöglichen und Informationen so zugänglich machen, dass sie den jeweiligen Anforderungen des Nutzers entsprechen.
- *Error prevention and correction*
Die Möglichkeit für Nutzerfehler sollte minimal sein und Fehleingaben durch den Nutzer rückgängig gemacht werden können.
- *User guidance and support*
Der Nutzer soll auch während der Durchführung Zugang zu Handbüchern und weiterer Unterstützung erhalten.

Alle Fragestellungen werden für jedes der neun Kriterien im Sinne der Auftretenshäufigkeit mit immer, meistens, manchmal oder nie spezifiziert und können mittels freier Kommentare vervollständigt werden. Ergänzend werden noch Probleme der Gebrauchstauglichkeit und deren Häufigkeit und Schweregrad sowie allgemeine Fragen zur Gebrauchstauglichkeit erfragt und festgehalten.

Die Methode an sich ist eine Evaluierungsmethode, die dann sinnvoll im Entwicklungsprozess eingesetzt werden kann, wenn ein Prototyp verfügbar ist. Hinsichtlich der Interpretierbarkeit ihrer Ergebnisse ist diese Methode nicht zum Vergleich unterschiedlicher Mensch-Maschine-Schnittstellen geeignet, da diese nur schwer quantifizierbar sind und verschiedene Probleme der Gebrauchstauglichkeit nur schwer gegeneinander aufzuwiegen sind. Dennoch wird angeraten, diese Methodik nach Möglichkeit mit Nutzerleistungsmessungen zu kombinieren, um herauszufinden, an welchen Stellen der Aufgabenbearbeitung Probleme der Gebrauchstauglichkeit konkret auftreten (Johnson, Clegg & Ravden, 1989). Nach Stanton et al. (2011) sind Checklisten leicht, kostengünstig und lediglich mit minimalem Training in unterschiedlichen Bereichen anwendbar und können flexibel an die durchzuführende Untersuchung angepasst werden. Dieses Vorgehen hat allerdings den Nachteil, kaum konsistent zu sein und rein auf den subjektiven Aussagen der Nutzer zu basieren. Für die Verwendung im Methoden-Inventar ist die Evaluation Checklist allerdings besonders gut in Situationen geeignet, in denen kein Vergleich verschiedener technischer Funktionen oder Situationen in der Interaktion mit der MMS Cockpit möglich ist (siehe Kapitel 4.1).

3.4.3.6 *Cognitive Walkthrough*

Der Cognitive Walkthrough als Methodik ist den Methoden der ‚Usability Inspection‘ zuzuordnen, einer Zusammenstellung an Methoden, die für die Identifikation von Problemen der Gebrauchstauglichkeit im Design, deren Schweregrad und zur Überprüfung der Anforderungen genutzt werden (Nielsen, 1994). Das Vorgehen des Cognitive Walkthrough wurde erstmals 1990 von C. Lewis, Polson, Wharton und Rieman eingeführt und beschrieben. Zunächst ist es notwendig, ein repräsentatives Aufgabenset zu identifizieren und für den Nutzer verfügbar zu machen – beispielsweise, aber nicht zwingend, mithilfe eines Simulators. Hierbei ist es notwendig, diejenigen Handlungsschritte des Nutzers zu identifizieren, die ein erfolgreiches Bearbeiten kennzeichnen. Zentral ist jedoch der Problemlöse- und Feedback-Evaluierungs-Prozess des Nutzers, der das Ziel hat, den Prozess zu kennzeichnen, wie der Nutzer das System erlernt. Mit Hilfe eines strukturierenden Sets an Fragen wird der Nutzer

erst an das Ziel seiner durchzuführenden Aufgabe herangeführt und die korrekte Durchführung der Einzelschritte zur Erlangung dieses Ziels hinsichtlich ihrer Einfachheit bewertet. Abschließend werden die Systemantworten beschrieben und es soll durch das Feedback des Nutzers festgestellt werden, ob der Nutzer die Ausgangsparameter für das Erreichen eines nachfolgenden Handlungsziels erkennt beziehungsweise ob er bemerkt, wenn die durchzuführende Aufgabe beendet ist. Anhand der Interaktionen des Nutzers mit der Schnittstelle werden dann die Entscheidungen, die im Designprozess getroffen wurden, explizit (C. Lewis et al., 1990).

Der Walkthrough kann also als Methode gesehen werden, in welcher der Nutzer das Aufgabenprofil durchläuft, Handlungen (mitunter nur zum Schein) durchführt und diese verbalisiert (Stanton et al., 2011). Als Methodik ist der Cognitive Walkthrough für das Methoden-Inventar geeignet, um die einzelnen Bedienungsschritte und somit die Schlüssigkeit der MMS hinsichtlich der Cockpit-Modifikation zu beurteilen. Dies liegt darin begründet, da die Methodik insbesondere auf Situationen ausgerichtet ist, in welchen der Nutzer keinerlei Handreichung durch vorbereitende Instruktionen erfährt, die Schnittstelle also selbsterklärend sein sollte (Hanington & Martin, 2012). Nachteilig ist, dass für die Anwendung des Walkthroughs primär, wenn nicht ausschließlich, Experten befragt werden müssen (Stanton et al., 2011).

3.4.3.7 Videokonfrontation

Durch die Verwendung von audiovisuellen Medien ist die Verhaltensbeobachtung ein zentraler Bestandteil der Daten-Erhebung und kann ergänzend zu subjektiven Berichten verwendet werden (Ellgring, 1995).

Die Videokonfrontation bezeichnet ein Verfahren, bei welchem der Operateur während der Durchführung seiner Nutzeraufgabe gefilmt und beobachtet wird und die Probleme, die während der Aufgabenbearbeitung entstanden sind, im Nachgang analysiert werden (Döhl, 2001). Die Nutzung einer Videokonfrontation zur Daten-Erhebung und -Klärung erfüllt den Zweck, die Situationen zielgerichtet herauszugreifen, die mit Hilfe eines Beobachtungsbogens und eines Zeitstempels während der Videoaufzeichnung insofern als relevant identifiziert wurden, als dass sie Gebrauchstauglichkeitsprobleme wie etwa Fehleingaben in den Bedienschritten, abweichende Handlungsziele des Nutzers oder allgemeine Aussagen zur Nutzung der technischen Funktion kennzeichnen. Im Rahmen einer Nachbesprechung werden dann die etwaigen Ziele und Zielkonflikte sowie das subjektive Handlungsvorgehen und die individuellen Gründe hierfür besprochen und festgehalten (Gediga, Greif, Monecke & Hamborg, 1989). Die Videokonfrontation eignet sich dazu, spezifische funktionsrelevante Benutzbarkeitsprobleme zu identifizieren, ist hierbei jedoch zeitaufwendig und kostenintensiv (Döhl, 2001).

Für das Methoden-Inventar ist die Videokonfrontation deshalb geeignet, da spezifische Benutzbarkeitsprobleme mit dieser Methode identifiziert werden können, ohne den Nutzer in der Bearbeitung seiner Aufgabe zu stören. Hinzu kommt, dass durch das hinweisgestützte Vorgehen vom Nutzer Informationen erfragt werden können, die ohne die videogestützten Hinweise so nicht mehr präsent gewesen wären.

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

3.4.4 Das multidimensionale Methoden-Inventar

Zur Bewertung der Qualität der MMS werden die in den Kapiteln 3.4.1, 3.4.2, und 3.4.3 aufgeführten Methoden zu einem Methoden-Inventar zusammengeführt und sukzessive in die Evaluierung von Cockpit-Modifikationen anhand der Einführung von neuen technischen Funktionen in den Cockpit-Simulator eingebunden. Die verwendeten Methoden sollen daher zum einen in die Simulationsumgebung integriert werden können und das Wissen der Piloten strukturiert erheben sowie deren Experten-Meinung kontrastieren. Zudem ist es eine Anforderung an das Methoden-Inventar, dass die eingesetzten Methoden für die Erhebung der Qualitätskriterien geeignet sind – also deren Konstrukte ausreichend erfassen. Tabelle 3.1 zeigt eine Übersicht der Methoden, die in den Evaluierungsexperimenten sukzessive zur Erhebung der Qualitätskriterien in der Simulationsumgebung eingebunden und verwendet worden sind.

Tabelle 3.1 Methoden-Inventar zur Erhebung der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit

<i>Methoden</i>	<i>Situationsbewusstsein</i>	<i>Workload</i>	<i>Gebrauchstauglichkeit</i>
	3D SART	NASA TLX	Nutzerleistung
	SAGAT	TDRT	PSSUQ
	Funktionsrelevante Maße	Eye-Tracking	Eye-Tracking
	Eye-Tracking		Interviews
			Videokonfrontation
			Cognitive Walkthrough
			Evaluation Checklist

Anmerkung: 3D SART = 3D Situation Awareness Rating Technique. SAGAT = Situation Awareness Global Assessment Technique. NASA TLX = NASA Task Load Index. TDRT = Tactile Detection Response Task. PSSUQ = Post Study System Usability Questionnaire. Die grauen Methoden sind qualitativ und können daher in der Validierung ab Kapitel 7 nicht berücksichtigt werden.

Zur Bestimmung des Situationsbewusstseins wird demnach der 3D SART als subjektiver Fragebogen verwendet, während die objektiven Maße des Eye-Trackings als indirekte Erhebungsform anzusehen sind. Die Verwendung des SAGATS als explizite Methode bietet die Möglichkeit, in einer angepassten Form sowohl retrospektiv als auch in Form von konkurrierenden Abfragen verwendet zu werden. Dies ist auch der Fall für situative, funktionsrelevante Maße. Wie Perrot et al. (2019) im Rahmen der Flugverkehrskontrolle anmerken, ist es ebenso notwendig, diejenigen Aspekte der Nutzer in objektive und messbare Kriterien umzuwandeln, die für die Operateure in der Bedienung einer Funktion oder eines Produktes als notwendig erachtet werden. Im Hinblick auf die Evaluierung einer Cockpit-Modifikation anhand einer konkreten technischen Funktion ist es daher möglich, dass Kriterien für das Situationsbewusstsein relevant werden, die zusätzlich erhoben werden müssen.

Die Workload der Nutzer wird anhand der subjektiven und multidimensionalen Rating-Skala NASA TLX und durch psycho-physiologische Maße des Eye-Trackings bestimmt. Außerdem soll die freie mentale Kapazität der Nutzer mittels einer taktilen DRT im Rahmen des Zweitaufgaben-Paradigmas erfasst werden.

3 Konstrukte und Methoden zur Qualitätsbestimmung der MMS Flugzeug-Cockpit

Die Nutzerleistung hingegen wird als klassisches Vorgehen zur Erhebung von Effektivität und Effizienz im Bereich der Gebrauchstauglichkeit angesiedelt. Hierunter fallen im Methoden-Inventar Bedienzeiten und Maße, die den Erfüllungsgrad der Nutzeraufgabe im Nutzungskontext erheben sowie Messungen, die die Genauigkeit erfassen, mit der die Nutzer diese Aufgaben bearbeiten. Die subjektiv empfundene Nutzerzufriedenheit wird anhand des PSSUQ erhoben. Als weiterhin objektives Messvorgehen werden Eye-Tracking-Maße verwendet, um Suchstrategien der Nutzer und die Dispersion von Displayinhalten zu kennzeichnen. Ergänzend werden qualitative Methoden wie Interviews, Videokonfrontation, der Cognitive Walkthrough und die Evaluation Checklist verwendet, um konkrete Aspekte des Designs der technischen Funktionen herauszuarbeiten und deren Auswirkungen auf die Nutzerzufriedenheit zu verdeutlichen.

4 INTEGRATION DES METHODEN-INVENTARS IN DIE SIMULATIONSUMGEBUNG

Da die Erhebung des Methoden-Inventars in einem komplexen sozio-technischen System umgesetzt wird, ist eine strukturierte Anpassung der Erhebungsmethoden an dieses System und die entsprechenden Domänenexperten notwendig. Dies impliziert, dass die Methoden nicht nur zuverlässig die Bewertungskriterien erheben sollen, sondern zusätzlich auch für die Versuchsumgebung und den Nutzerkontext geeignet sein müssen (vgl. Kapitel 2.4). In zwei Vorstudien konnte ein Verfahren entwickelt werden, welches die Implementierung des multidimensionalen Methoden-Inventars in eine Versuchsumgebung zur Evaluierung einer Mensch-Maschine-Schnittstelle (MMS) mit einem Cockpit-Simulator erlaubt. Dieser Abschnitt widmet sich der Darstellung der prozeduralen Integration des multidimensionalen Methoden-Inventars in die Simulationsumgebung. Der Prozess beginnt mit einer detaillierten Analyse und Charakterisierung der technischen Cockpit-Modifikation, des möglicherweise vorhandenen Designs sowie des geplanten Cockpit-Designs. Ergänzend muss die Aufgabe, in welcher die entwickelte technische Funktionalität genutzt werden soll, strukturiert erfasst werden. Hierzu wird ein repräsentativer Anwendungsfall ausgewählt und die Nutzung der technischen Funktionalität in diesem Anwendungsfall strukturiert und logisch beschrieben. Anhand dieser Beschreibung lassen sich die Messmethoden anpassen sowie die zu verwendenden Messzeitpunkte ableiten. Mit Hilfe dieser Strategie kann ein Testszenario als experimentelles Vorgehen festgelegt werden.

Eine Evaluierung in einer Simulationsumgebung findet in dem Spannungsfeld zwischen Feld- und Laborstudie statt. Je nach Ziel der Evaluierung kann es angebracht sein, die Untersuchung als Feld- oder als Laborstudie zu definieren. Eine Feldstudie wäre beispielsweise dann angebracht, wenn die Cockpit-Modifikation so ausgereift ist, dass diese kurz vor der Auslieferung steht und an den Nutzer übergeben werden soll. In diesem Fall sind also weitgehende Funktionsänderungen nicht mehr zu erwarten. Hier wird die Simulationsumgebung so realistisch wie möglich ausgelegt, um dem Experten einzuräumen, die Cockpit-Modifikation hinsichtlich der Erfüllung der Anforderungen zu bewerten.

Die Durchführung der Nutzer-Evaluierungen in einem frühen Stadium des Entwicklungsprozesses orientieren sich aber eher an Laborstudien. Das heißt, dass eine Evaluierung der Cockpit-Modifikation mit Hilfe der Qualitätskriterien in einem Spannungsfeld zwischen einem möglichst repräsentativen Nutzungskontext und dem zielgerichteten – auch intrusiven – Einsatz des Methoden-Inventars stattfindet. Eine gewisse Labornähe ist notwendig, um belastbare Aussagen treffen zu können. Das heißt, um in einer künstlichen, jedoch anwendungsnahen Situation messbare Ergebnisse zu erhalten, ist es notwendig, dass die Situation selbst über alle Probanden hinweg vergleichbar ist. Das heißt, dass alle Probanden zum gleichen Zeitpunkt und im gleichen situativen Kontext die identischen Informationen zur Verfügung haben müssen und die gleichen Aufgaben tätigen. Hierzu ist es zum einen erforderlich, eine entsprechende Vergleichsbasis zu schaffen (siehe 4.1) und zum anderen muss der Handlungskontext so strukturiert erfasst werden, dass eine zielgerichtete Implementierung der Methoden und deren Messzeitpunkte (siehe 4.2) möglich ist. Erst dann kann das Evaluierungsvorhaben in die Simu-

lationsumgebung integriert werden (siehe 4.3) und das Design der technischen Cockpit-Funktion von einer Experten-Stichprobe überprüft und verifiziert (siehe 5) werden.

4.1 Experimentelle Erhebung in der Cockpit-Simulation

Zur experimentellen Erhebung in einem Cockpit-Simulator ist in einem ersten Schritt festzulegen, welche Aussagen hinsichtlich des Designs der technischen Funktion getroffen werden sollen und welche Vergleichsform zur Überprüfung der Qualitätskriterien notwendig ist.

Eine mögliche Vergleichsform wäre, zwei oder mehrere Designs von Cockpit-Modifikationen miteinander zu vergleichen. Beispielsweise könnte ein neues Design mit dem bisherigen, bestehenden Design verglichen werden. So ist festzustellen, ob das neue Design den Nutzer in seiner Aufgabe tatsächlich besser unterstützt und nicht zu übermäßigen zusätzlichen Belastungen führt oder das Aufrechterhalten des Situationsbewusstseins des Operateurs einschränkt. Eine andere Variante dieser Vergleichsform wäre es, zwei verschiedene Designs einer neu zu entwickelnden Funktion zu vergleichen, um herauszufinden, welcher Designvorschlag für den Operateur besser nutzbar ist.

Beide Vorgehensweisen des Designvergleichs sind in ihrer Anwendung zielgerichtet, der Aufwand für die Umsetzung in der Simulationsumgebung ist aber relativ hoch, da jeweils zwei Softwarestände, Aufgabenanalysen und vergleichbare Simulations-Szenarien entwickelt werden müssen. Es empfiehlt sich daher, die Nutzeraufgabe, in welcher die Funktion getestet werden soll, sehr eng zu umgrenzen. Je nach Umfang und Art der verwendeten Methoden kann hier auch ein hoher Erhebungs- und Auswerteaufwand entstehen. Daher sollten nach Möglichkeit automatisierte Messungen wie beispielsweise die Erhebung von Bedienzeiten mittels Cockpit-Simulators verwendet und vorgenommen werden, zumal dieses Vorgehen auch weniger intrusiv in der Anwendung ist. Da bei diesen Vergleichsformen sowohl objektive als auch subjektive Methoden so eingesetzt werden können, dass sie zu einer spezifischen Aussage hinsichtlich einer objektiven Designpräferenz führen können, lohnt es sich allerdings, dieses Vorgehen wann immer möglich zu wählen.

Eine weitere Option ist es, das Design einer technischen Funktion in verschiedenen situativen Nutzerkontexten zu vergleichen. Im Rahmen eines solchen Vergleichs ist es möglich, die neue Cockpit-Modifikation unter verschiedenen Umwelt- oder Situations-Einflüssen wie beispielsweise unter Einfluss von Vigilanz, einer Veränderung des Lärmpegels, Aufgabenwichtigkeit, Zeitdruck oder ähnlichem zu untersuchen. Je nach Komplexität der Situation gestaltet sich aber eine solche Erhebung als schwer kontrollierbar, da immer noch sichergestellt werden muss, dass die Nutzer zur gleichen Zeit die gleichen Handlungen tätigen und die gleiche Information erhalten, um zu einem messbaren Ergebnis zu gelangen. Das heißt im Rückschluss auch, dass sowohl die Nutzeraufgabe als auch die Situationen detailliert beschrieben werden müssen und auch deren Vergleichbarkeit sichergestellt sein muss, indem die Variation der Situation auf den zu variierenden Faktor beschränkt wird. Mit Hilfe dieses Vorgehens ist es jedoch möglich, die Performanz eines Designs detaillierter, wenn auch nicht umfassend zu betrachten. Es empfiehlt sich daher, für den Vergleich möglichst repräsentative Situationen zu wählen.

4 Integration des Methoden-Inventars in die Simulationsumgebung

Weiterhin ist auch eine Kombination der oben genannten Vergleiche möglich. Das heißt, ein Design kann auch innerhalb einer Erhebung mit verschiedenen anderen Designs derselben Funktion und in unterschiedlichen situativen Kontexten, beispielsweise durch Einführung von verschiedenen Belastungsstufen, überprüft werden. Da sich in diesem Fall der Aufwand schnell potenziert, ist es jedoch zwingend notwendig, die Testsituation sehr stark einzuschränken und notfalls auch auf eine artifizielle laborähnliche Untersuchung zurückzugreifen. Andererseits können gerade hier subjektive und objektive Messmethoden so eingesetzt werden, dass eine umfangreiche Bewertung des zu untersuchenden Designs und dessen Performanz im Hinblick auf die Bedienung durch den Nutzer möglich ist.

Wenn stattdessen mehrere unterschiedliche Funktionen in einem Vorgehen überprüft werden sollen, ist auf Grund der Verschiedenheit der jeweiligen Situationen und Aufgaben, in welchen die Cockpit-Modifikationen bedient werden, kein Vergleich im herkömmlichen Sinn möglich. Das heißt, dass dann letztendlich nur die subjektiven Aussagen der Nutzer zu einer absoluten Bewertung der jeweiligen Cockpit-Modifikationen genutzt werden können. Damit deren Aussagen für eine weitere Verfeinerung des Designs verwendet werden können, ist es aber auch hier erforderlich, ein repräsentatives, situatives Umfeld für die einzelnen Funktionen zu schaffen. Mit strukturierten Befragungen der Experten ist es zwar möglich, deren subjektive Designpräferenzen und auch interessierende Bereiche für Anpassungen der Cockpit-Modifikationen zu identifizieren; vergleichende Aussagen, die zu einer objektiven Bewertung des Designs oder zu spezifischeren Aussagen hinsichtlich der Performanz des Designs in verschiedenen Situationen führen, allerdings nicht. Somit können zwar mehrere Funktionen unter Verwendung von qualitativen Methoden, die in diesem Fall schnell eingesetzt werden können, betrachtet werden; das Treffen von konkreten Aussagen ist jedoch nicht möglich, da weder die Einführung des Designs noch der situative Kontext des Nutzers kontrolliert werden.

Eine Übersicht der verschiedenen Vergleichsformen sowie eine Zusammenfassung von deren Vor- und Nachteilen ist in Tabelle 4.1 zu finden.

Tabelle 4.1 Gegenüberstellung der Vor- und Nachteile experimenteller Vergleiche in der Simulationsumgebung

Experimenteller Vergleich	Vorteile	Nachteile
Vergleich zweier oder mehrerer Designs	<ul style="list-style-type: none"> • Einsatz objektiver und subjektiver Methoden • Objektive Empfehlung eines Designs 	<ul style="list-style-type: none"> • Erhöhter Design- und Simulationsaufwand • Hoher Erhebungs- und Auswertungsaufwand
Vergleich des Designs in zwei oder mehreren verschiedenen Situationen	<ul style="list-style-type: none"> • Einsatz objektiver und subjektiver Methoden • Ermöglicht Aussagen über das Design in dessen situativer Anwendung 	<ul style="list-style-type: none"> • Hoher Planungsaufwand bei Detaillierung der Nutzeraufgabe • Hoher Erhebungs- und Auswertungsaufwand • Schwer kontrollierbar
Vergleich von Designalternativen in verschiedenen Situationen	<ul style="list-style-type: none"> • Einsatz objektiver und subjektiver Methoden • Objektive Empfehlung eines Designs • Ermöglicht Aussagen über die untersuchten Designs in ihrer situativen Anwendung • Stark kontrolliertes Vorgehen 	<ul style="list-style-type: none"> • Erhöhter Designaufwand • Artificielle Datenerhebung • Nur in einem eng umgrenzten situativen Kontext möglich • Hoher Erhebungs- und Auswertungsaufwand
Befragung ohne Vergleich	<ul style="list-style-type: none"> • Schnelle und zielgerichtete Anwendung und Auswertung qualitativer Methoden • Zielgerichtete Erfassung von subjektiven Expertenaussagen • Überprüfung verschiedener Funktionen 	<ul style="list-style-type: none"> • Nicht kontrollierbar • Nur qualitative Befragung der Experten • Keine objektiven Aussagen zu Designpräferenzen oder situativer Performanz

Sofern definiert wurde, welche Vergleichsform zur Evaluierung der Cockpit-Modifikation genutzt werden soll, muss festgelegt werden, wie und zu welchen Zeitpunkten im Experiment die Cockpit-Modifikation konkret überprüft werden soll und kann. Dies ist insbesondere für die ersten drei Vergleichsformen notwendig, da hier eine besonders genaue Vorstellung der Handlungen im Cockpit unter Anwendung der Cockpit-Modifikation und des Nutzerkontextes in Form einer repräsentativen Situation vorliegen muss. Dies erlaubt dann einen zielgerichteten Einsatz des Methoden-Inventars, da so die Messzeitpunkte offenbar werden (siehe Kapitel 4.2). Zudem muss die Vergleichbarkeit auch im Simulations-Geschehen sichergestellt werden (siehe Kapitel 4.3) – das heißt, dass alle Experten zum gleichen Zeitpunkt in der Simulation die gleichen Handlungen ausführen und die gleichen Informationen zur Verfügung haben.

4.2 Entwicklung von Messbedingungen im situativen Kontext

Bei der Evaluation von Cockpit-Modifikationen ist eine Voraussetzung für Vergleichbarkeit und daher auch Voraussetzung für den zielgerichteten Einsatz des Methoden-Inventars, dass die Handlungen im Cockpit strukturiert erfasst, die Ziele des Nutzers deutlich werden und in einen eindeutig definierten zeitlichen Handlungsablauf überführt werden. Das heißt, dass Nutzungskontext und Nutzerverhalten im Hinblick auf die Cockpit-Modifikation so detailliert beschrieben werden müssen, dass auch ersichtlich wird, wann die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit für den Nutzer relevant werden. Dieses Vorgehen erlaubt eine Festlegung der Messzeitpunkte und ermöglicht somit auch die Methodenauswahl. Demnach wird definiert, wann gemessen wird, was gemessen wird und wann die Messung zu beenden ist. Dies hat darüber hinaus den Vorteil, dass die Komplexität der Handlungen im Cockpit reduziert wird und auch für den evaluierenden Versuchsleiter ersichtlich wird.

Zur strukturierten Erfassung von Nutzerkontext und Nutzungsverhalten im konkreten Anwendungsfall ist eine Erarbeitung folgender Aspekte notwendig:

1. Charakterisierung des technischen Systems
2. Erhebung und Beschreibung eines repräsentativen Anwendungsfalls
3. Erstellung einer Aufgabenanalyse zur Erfassung des Nutzerverhaltens
4. Erstellung eines Handlungsplans unter Kontrastierung der zu treffenden Entscheidungen

In einem ersten Schritt ist es notwendig, die technische Auslegung der Cockpit-Modifikation, deren Design und deren Bedienung zu erfassen. Hierzu kann auf die dokumentierten Anforderungen und auf vorhandene technische Handbücher und Dokumente zurückgegriffen werden, die beschreiben, welche Bedienelemente und welche Bedienschritte innerhalb der Menüführung notwendig sind.

Als nächstes wird bestimmt, in welchen typischen Situationen die Bedienung der Cockpit-Modifikation stattfindet und welche Ziele der Nutzer während der Bedienung der technischen Funktion verfolgt. Es ist an dieser Stelle notwendig, den Nutzer einzubinden und durch Experteninterviews dessen Wissen abzufragen. Ein repräsentativer Anwendungsfall wird auf dieser Grundlage identifiziert und ausformuliert. Unter einem repräsentativen Anwendungsfall, dem sogenannten Use Case, versteht man die Beschreibung des Systemverhaltens unter verschiedenen Bedingungen und in Abhängigkeit der Nutzeranfragen und Interaktionen mit dem System, die der Anwender tätigt, um ein Ziel zu erreichen (Cockburn, 2001). Der Anwendungsfall als solches ist in eine typische, repräsentative Handlungssituation eingebettet bzw. tritt darin auf. Das heißt, dass der Nutzer die technische Cockpit-Funktion in einer typischen Art und Weise nutzt und diese in der Regel in einer repräsentativen Situation – dem Nutzungskontext – bedient. Diese Situation bildet schließlich das Szenario aus, in welcher der Anwendungsfall der technischen Funktion getestet werden kann und kann als die Umweltbedingungen verstanden werden, in welcher der Anwendungsfall auftritt (siehe Kapitel 4.3).

Ist ein repräsentativer Anwendungsfall festgelegt und wurde bestimmt, welche Begleitumstände das Szenario bilden, kann damit begonnen werden, das Szenario und den Anwendungsfall struktu-

4 Integration des Methoden-Inventars in die Simulationsumgebung

riert und somit so vollständig und detailliert wie notwendig in einem weiteren Schritt zu erfassen. Dies ist grundlegend für die Implementierung von Messmethoden in ein komplexes und sozio-technisches System, da mit strukturierenden Methoden wie beispielsweise Aufgabenanalysen das zu untersuchende System im Nutzungskontext beschrieben werden kann und mögliche Messzeitpunkte identifiziert werden können. Dies ist möglich, da Aufgabenanalysen dazu genutzt werden, die Mensch-Mensch-Interaktionen ebenso zu verstehen wie die Mensch-Maschine-Interaktionen. Das erfasste Expertenwissen wird dazu genutzt, um die notwendigen individuellen Aufgabenschritte und Interaktionen, die hierfür physischen Handlungen und die genutzte Technologie zu detaillieren und diese in eine zeitliche Abfolge zu überführen (Stanton et al., 2011). Eine spezifische und häufig verwendete Form der Aufgabenanalyse ist die Hierarchische Task-Analyse (HTA), die von Annett und Duncan (1967) mit entwickelt wurde. Die HTA wird verwendet, um die Systemfaktoren zu verstehen, die menschliche Leistung beeinflussen (Shepherd, 2005). Aufgaben werden hierzu in Form von hierarchischen (Nutzer-) Zielen beschrieben und diese werden so lange in Unterziele unterteilt, bis die Handlungsebene des Nutzers erreicht wird (Stanton et al., 2011). Die rein logische und hierarchische – nicht aber zeitliche – Struktur wird anschließend reformuliert und in einen Plan übersetzt. Mit der Überführung in einen Plan wird dann der sequentielle Ablauf der Nutzerhandlungen und dessen Entscheidungen deutlich (Annett, 2003). Die HTA hat somit den Vorteil, dass sie beschreibt, welche Ziele der Nutzer erreichen möchte, welche Entscheidungen er trifft und welche kognitiven und physischen Tätigkeiten er hierzu ausübt.

Die Erstellung einer HTA ist Teil und Grundlage für die Anwendung eines multidimensionalen Methoden-Inventars innerhalb eines Cockpit-Simulators, da die Aufgaben, die der Nutzer zu erfüllen hat, in der Realität in ein komplexes sozio-technisches System eingebunden sind, in dem der Nutzer seine Ziele als Teil einer Mission verwirklicht. Diese Komplexität muss so operationalisiert werden, damit Messpunkte zielgerichtet in die zu evaluierende Situation eingebaut werden können und so das Nutzerverhalten in der Messumgebung untersucht werden kann.

Zur Einbindung von Messungen des Situationsbewusstseins muss bekannt sein, an welchen Stellen im Handlungsplan das Situationsbewusstsein offenbar wird. Die Erhebung von Situationsbewusstsein erfolgt anhand der zu treffenden Entscheidungen des Nutzers, da hier Ausmaß und Bedeutung der Informationen, die der Nutzer zum Treffen dieser Entscheidung benötigt, bekannt sein müssen (Endsley, 2000a). Um nun zielgerichtet Messzeitpunkte für die Erhebung von Situationsbewusstsein zu identifizieren, ist die Einführung einer Situation Awareness Requirements Analyse (SARA) nach Endsley (1993) geeignet.

Eine beispielhafte Verwirklichung eines HTA-Handlungsplans und einer SARA zeigen Morgenstern et al. (2016) an der Aufgabenstellung ‚Flugzeugsteuerung‘ auf, wie sie in Abbildung 4.1 dargestellt wird.

4 Integration des Methoden-Inventars in die Simulationsumgebung

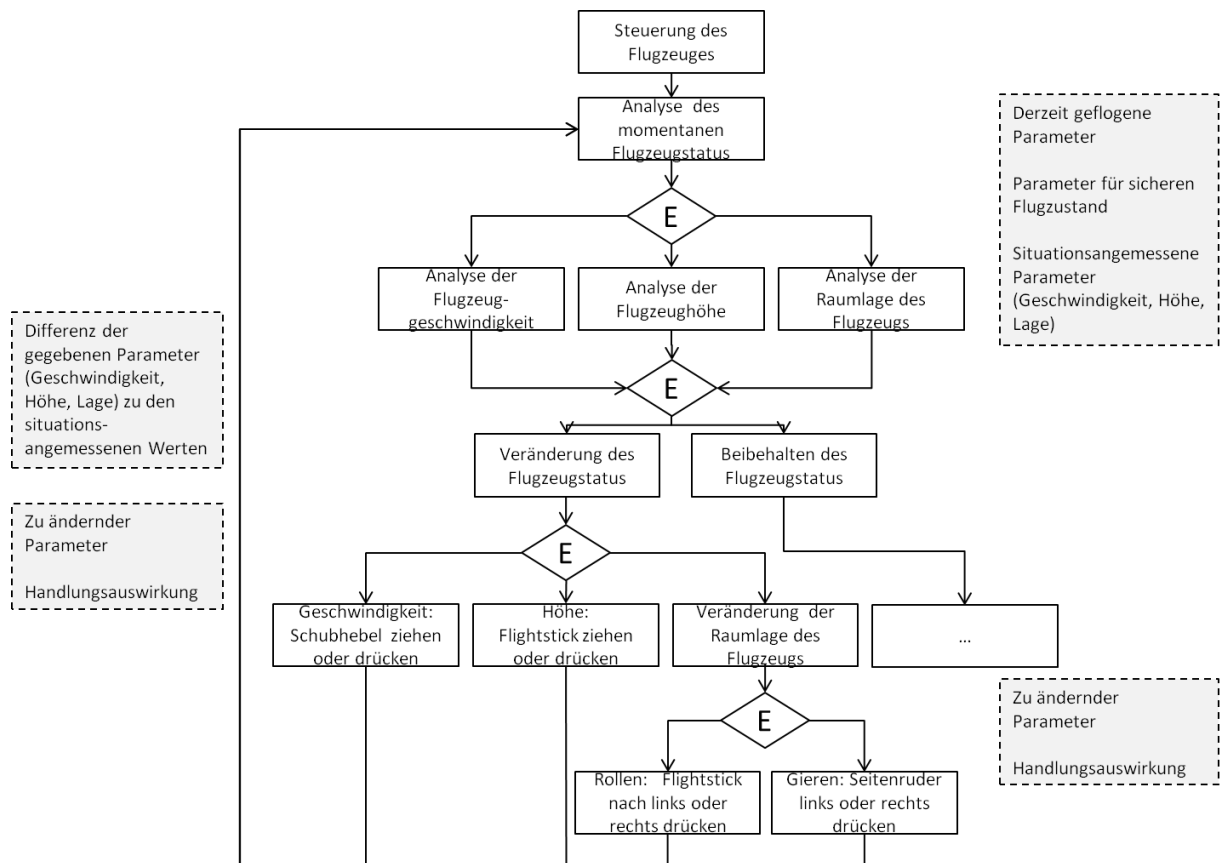


Abbildung 4.1 Exemplarischer HTA-Plan der Aufgabe ‚Flugzeugsteuerung‘ mit integrierter SARA nach Morgenstern et al. (2016)

Auf Grundlage einer HTA zur Flugzeugsteuerung nach Kellerer (2010) wurde der Handlungsplan von Morgenstern et al. (2016) ausformuliert und mit einer SARA ergänzt. Die Entscheidungspunkte des Piloten werden hier expliziert und sind an den mit ‚E‘ gekennzeichneten Rauten zu erkennen. In den gestrichelten, grauen Boxen sind jeweils die Informationen aufgeführt, die dem Piloten bekannt sein müssen, ehe dieser eine Entscheidung treffen kann wie beispielsweise die Entscheidung zur Veränderung des momentanen Flugzeugstatus. Mit Hilfe des Handlungsplanes wird somit ersichtlich, wie der Pilot seine Ziele erreicht (s.a. Stanton et al., 2011). Mittels SARA wird das mentale Modell des Piloten hinsichtlich Zeitpunkt, Aufgabe und Ziel explizit und die Anforderungen an das Situationsbewusstsein werden deutlich (Morgenstern et al., 2016).

Ergänzend zu den aufgeführten strukturierenden Vorgehensweisen können die Priorisierungsstrategien des Piloten mittels Eisenhower-Matrix abgebildet werden. Die Eisenhower-Matrix wird üblicherweise im Bereich des Zeitmanagements verwendet und unterteilt anhand einer Vier-Felder-Tafel vorhandene Aufgaben nach Wichtigkeit und Dringlichkeit (s.a. Hering, 2014). Zur Kontrastierung gradueller Unterschiede in der Informationssuche ist es demnach möglich, Flugaufgaben nach Wichtigkeit und Dringlichkeit zu unterteilen. In diesem Zusammenhang ist Wichtigkeit als wertbasierte Aufgabenwichtigkeit zu verstehen, welche die Notwendigkeit in sich birgt, eine derzeit ausgeführte Aufgabe zu unterbrechen und sich einer anderen Aufgabe zuzuwenden. Dringlichkeit hingegen wirkt sich lediglich auf den vom Piloten empfundenen Zeitdruck aus, eine bestimmte Aufgabe auszuführen.

4 Integration des Methoden-Inventars in die Simulationsumgebung

Mit Anwendung dieses Vorgehens ist es möglich, einen situativen Vergleich zur Evaluierung einer technischen Funktion im Cockpit-Simulator, aufbauend auf den zuvor genannten strukturierenden Verfahren, vorzunehmen.

Zur direkten Implementierung des Handlungsplanes in den situativen Nutzerkontext innerhalb der Simulationsumgebung sind weitere Maßnahmen notwendig, wie Kapitel 4.3 zeigt.

4.3 Umsetzung des Methoden-Inventars in der Simulationsumgebung

Um die Bedienung der technischen Funktion und den zugehörigen Handlungsplan zielgerichtet in die Simulationsumgebung zu integrieren, ist es erforderlich, den repräsentativen Anwendungsfall in einen situativen Handlungsrahmen – das Szenario – einzubetten. Hierzu wird die Anwendung der technischen Funktion bzw. der konkrete Anwendungsfall als Ausgangspunkt genommen und in einen Handlungskontext eingebunden, in der diese Funktion (nach Möglichkeit auch mehrmals) genutzt werden kann. Der Handlungskontext wird ebenfalls anhand von Experteninterviews genau definiert und bildet das gesamte situative Geschehen während eines Szenarios ab. So sind neben dem aktuellen Auslieferungszustand des Flugzeuges, dessen Ausrüstung und Konfiguration auch die Umweltbedingungen, in denen die Bedienung der technischen Funktion stattfinden soll, festzulegen.

Weiterhin sind alle notwendigen Akteure als Teil des Szenarios zu definieren, um den Piloten während der Simulation ein realistisches Situations-Empfinden zu vermitteln. Diese Akteure können sowohl artifiziell sein, aber auch als Realpersonen in das Situations-Geschehen eingebunden sein. Daher ist auch die Erarbeitung eines Drehbuches erforderlich. Hierin wird neben dem zeitlichen Ablauf und den Handlungen der Akteure auch die standardisierte Kommunikation festgehalten, um den Piloten eine Immersion in das Situations-Geschehen zu ermöglichen. Es empfiehlt sich somit, mit einem Experten bei der Erstellung eines Szenarios eng zusammenzuarbeiten und wenn möglich und notwendig, diesen auch als Konföderierten in das Situations-Geschehen einzubinden.

Sobald die Handlungen und Ereignisse im Cockpit eindeutig definiert und in eine stringente Zeitlinie überführt werden können, kann das Szenario fixiert und eine Vergleichbarkeit über alle Probanden hinweg sichergestellt werden. Alle Probanden müssen hierzu die gleichen Handlungen zum gleichen Zeitpunkt im Simulations-Geschehen tätigen und die gleichen Informationen zur Verfügung haben. Eine Möglichkeit, um dies sicherzustellen, ist es, die jeweiligen Auslöseereignisse für die zu untersuchende Handlungen festzuhalten und zu kontrollieren. Auslöseereignisse können beispielsweise in Form von verbalen Ankern im Drehbuch auftreten, aber auch situative Ereignisse sein, die zeitlich und räumlich im Szenario festzuhalten sind.

Weiterhin gilt es, die Rollen der Akteure, aber auch die des Personals während des Evaluations-Geschehens im Cockpit-Simulator eindeutig vorzugeben und zuvor zu trainieren. So sind auch die Aufgaben, welche die Versuchsleiter während der Evaluierung durchführen, ebenso klar zu definieren wie die Aufgaben innerhalb des Simulations-Geschehens. Hierzu gehören Aufgaben wie das Beobachten des zu untersuchenden Verhaltens, das Setzen zeitlicher Markierungen während des Simulations-Geschehens zur Nachverfolgung im De-briefing, aber auch der Einsatz sämtlicher Methoden

4 Integration des Methoden-Inventars in die Simulationsumgebung

und deren Überwachung. Ein Beobachter sollte hierbei möglichst wenige Verhaltensweisen gleichzeitig beobachten, markieren und kommentieren. Aufgrund der Komplexität von Evaluierungen in einer Simulationsumgebung empfiehlt sich die Durchführung einer Probeevaluierung. Dies dient dem Training des Personals, stellt sicher, dass die Evaluierung dem gewünschten zeitlichen Ablauf folgt und sowohl Evaluierungsgegenstand als auch methodisches Vorgehen zielgerichtet in einer funktionalen Simulationsumgebung ein- und umgesetzt werden können.

Zusammenfassend lässt sich daher festhalten, dass aufgrund der Vielfalt der Ereignisse im Cockpit und der somit erschwerten Kontrollierbarkeit der Handlungen während der Evaluierung auf ein eng umgrenztes und für die Versuchsleiter handhabbares Szenario geachtet werden sollte. So ist es möglich, die Vergleichbarkeit sicherzustellen, die zur Untersuchung der Bedienung einer technischen Funktion im situativen Handlungskontext mit Hilfe eines Cockpit-Simulators erforderlich ist.

5 DIE EVALUIERUNG DES MULTIDIMENSIONALEN METHODEN-INVENTARS

Mit der Evaluierung des multidimensionalen Methoden-Inventars kann begonnen werden, wenn dieses erfolgreich in die Simulationsumgebung integriert ist (siehe Kapitel 4) und die hierfür notwendigen repräsentativen Untersuchungsszenarien festgelegt sind. Im Folgenden werden die Untersuchungen von vier verschiedenen Cockpit-Modifikationen vorgestellt. Die Untersuchungen entsprechen den ersten drei der unter Kapitel 4.1 erklärten Vergleichsformen in der Cockpit-Simulation:

- Vergleich zweier oder mehrerer Designs
- Vergleich eines Designs in (mindestens) zwei Situationen
- Vergleich von Design-Alternativen in verschiedenen Situationen (beispielsweise Belastung)
- Befragung ohne Vergleich

Da die Befragung ohne Vergleich nicht zur Validierung des Methoden-Inventars beiträgt, weil nur eine Befragung anhand von Interviews oder strukturierten Checklisten für eine genauere Betrachtung der verschiedenen Funktionalitäten möglich ist, ist diese Evaluierungsform nicht weiter von Belang. Stattdessen wird der Vergleich zweier Designs umfänglich für die Validierung verwendet, da mit der ‚Machbarkeitsstudie zur Implementierung einer Navigationshilfe‘ (Kapitel 5.1) ein neues Design mit der ursprünglichen Cockpit-Auslegung verglichen wird. Dies wird ergänzt durch die ‚Untersuchung der Bedienung eines Sensors zur Einstellung von Genauigkeit und Sensitivität‘ (Kapitel 5.4), in welcher zwei verschiedene Designalternativen evaluiert werden. Die ‚Untersuchung des Re-Designs einer Navigationshilfe‘ (Kapitel 5.2) baut direkt auf der Machbarkeitsstudie auf und analysiert die gleiche Funktion nach einem Re-Design in zwei verschiedenen komplexen Situationen. Bereits 2016 haben Rabl et al. eine Studie zur Untersuchung eines räumlich-auditiven Warnsignals durchgeführt, in welcher das Methoden-Inventar dazu verwendet wurde, eine neue Cockpit-Funktionalität mit der bisherigen Cockpit-Auslegung und unter Einführung von verschiedenen Belastungsstufen (Kapitel 5.3) zu untersuchen.

Da die Evaluierung in der Simulationsumgebung in allen vier Funktionsuntersuchungen sich sehr ähnelt und der Schwerpunkt dieser Arbeit auf der Validierungsstrategie liegt, wird nur ein Evaluierungsbeispiel für die Darstellung der Datengrundlage beschrieben. Hierzu wird die ‚Untersuchung des Re-Designs einer Navigationshilfe‘ (Kapitel 5.2) gewählt und insbesondere hinsichtlich der ‚Bewertung der Verwendung des Re-Designs einer Navigationshilfe‘ unter Betrachtung der Qualitätskriterien detailliert, weil in dieser Studie das Methoden-Inventar umfassend eingesetzt wird. Die Untersuchungen der weiteren Funktionalitäten werden daraufhin beschrieben.

Des Weiteren wird im Hinblick auf alle Evaluierungen eine integrierte Beschreibungsform gewählt, welche sich an dem ursprünglichen Adressatenkreis an Human-Machine-Interface (HMI)-Designern orientiert und daher auch Bestandteile des Szenarios, des Versuchsablaufs sowie eine integrierte Bewertung der Cockpit-Modifikationen unter Einsatz des verwendeten Methoden-Inventars enthält. Abschließend wird auch die jeweils eingesetzte Methoden-Kombination hinsichtlich des praktischen Einsatzes betrachtet und bewertet. In diesem Kapitel soll somit verdeutlicht werden, wie und anhand

welcher Cockpit-Funktionalitäten die Evaluierungen vorgenommen wurden (siehe Kapitel 5.1, 5.2, 5.3, 5.4). Insbesondere in Kapitel 5.2 liegt allerdings der Fokus auf dem Methoden-Einsatz und die darauf aufbauende Datenbasis, die für die spätere Validierungsstrategie genutzt wird und weniger auf der inhaltlichen Bewertung der Cockpit-Modifikationen, was zur Bewertung des eingesetzten Methoden-Inventars eine integrierte Schreibform erfordert.

5.1 Machbarkeitsstudie zur Implementierung einer Navigationshilfe

In die bestehende Auslegung des Flugzeuges ist eine neue Funktion integriert worden, welche es dem Piloten ermöglicht, geografische Hilfsziele in der Luft, sogenannte Weg- oder Navigationspunkte, anhand von einem feststehenden, fünfstelligen Identifizierungs-Code einzugeben. Bislang war es den Piloten nur möglich, einen Wegpunkt anhand von geografischen Koordinaten einzugeben, die vom Piloten entweder zuerst in einem Tabellenwerk nachgeschlagen oder mit Hilfe einer Karte gesucht werden, bevor sie in das System eingetippt werden können. Dieser Vorgang ist in der Regel zeitaufwendig und fehleranfällig.

In einem Experiment ist daher untersucht worden, ob mit Einführung einer Navigationshilfe die Eingabe von Wegpunkten gebrauchstauglicher gestaltet, die Arbeitsbelastung vermindert und das Situationsbewusstsein des Piloten gestärkt werden kann.

5.1.1 Stichprobenbeschreibung

Es sind vier männliche Domänenexperten (siehe Kapitel 2.2) im Alter von 40, 39, 37 und 46 Jahren mit unterschiedlicher Flugerfahrung untersucht worden. Alle Probanden haben freiwillig an der Untersuchung teilgenommen und sind hinsichtlich der anonymisierten Weiterverarbeitung ihrer Daten zu Forschungszwecken aufgeklärt worden.

5.1.2 Verwendete Methoden

Auf Grundlage einer hierarchischen Task-Analyse sind ein Szenario und ein zugehöriges Drehbuch entwickelt worden, die für eine möglichst repräsentative Erhebung des Szenarios sorgen. Das Qualitätskriterium Situationsbewusstsein wurde anhand der Erhebung von Blickdaten gemessen. Neben dem Blickverhalten ist hinsichtlich der Bestimmung von Workload auch der NASA Task Load Index (NASA TLX) verwendet worden. Die Gebrauchstauglichkeit wurde mit Hilfe der Bedienzeiten und der Erfassung der Blickdaten objektiv bewertet und die Nutzerzufriedenheit wurde subjektiv mit Hilfe des Post Study System Usability Questionnaire (PSSUQ) und ergänzend mit halbstrukturierten Interviews erfasst. Eine Übersicht der verwendeten Methoden aus dem Methoden-Inventar ist in Tabelle 5.1 nachzulesen.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Table 5.1 Methoden-Inventar zur Erhebung der Qualitätskriterien in der Machbarkeitsstudie

Methoden	Situationsbewusstsein	Workload	Gebrauchstauglichkeit
	Eye-Tacking	NASA TLX Eye-Tracking	Nutzerleistung PSSUQ Eye-Tracking Interviews Videokonfrontation

Anmerkung: Die qualitativen Methoden (Interviews und Videokonfrontation) sind nicht Teil der ab Kapitel 6 beschriebenen Validierung

5.1.3 Szenario-Beschreibung

Die Versuchspersonen sollten mit Hilfe des Cockpit-Simulators nach einer zuvor geplanten Route vom Flugplatz Nörvenich bei schlechten Wetterverhältnissen zum Flugplatz Manching fliegen. Während des Fluges wurden die Versuchspersonen zwei Mal via Sprechfunk angewiesen, eine Umleitung, bestehend aus jeweils vier Wegpunkten, abzufliegen. Aufgabe der Versuchspersonen war es, in einem Szenario alle acht Wegpunkte einmal mit dem bestehenden Cockpit-Design einzugeben und im zweiten Szenario, auf einer ähnlichen Flug-Route, mit der Navigationshilfe. Hierbei wurde zudem bei der bestehenden Cockpit-Auslegung nach den Eingabeformen ‚Tabellenwerk‘ und ‚Karte‘ unterschieden und in der Bedingung, in welcher die Navigationshilfe eingeführt wurde, nach zwei verschiedenen Bedienkonzepten.

5.1.4 Versuchsablauf

Nach Begrüßung der Probanden und der Einholung ihrer Einverständniserklärung hinsichtlich der Verarbeitung ihrer Daten wurden sie über Ziel und Zweck der Evaluierung aufgeklärt. Anschließend erhielten sie eine Einweisung in die Bedienung der neuen Navigationshilfe, bevor die Piloten in der Simulationsumgebung die zu untersuchenden Bedienungen trainieren konnten. Nach dem Training erhielten die Probanden ein Briefing hinsichtlich des Szenario-Ablaufes und erhielten die hierfür notwendigen Informationen wie beispielsweise der Füllstand des Tanks, vorherrschende Wetterverhältnisse und die Position des Flugzeuges bei Simulationsstart. Danach wurde das Eye-Tracking kalibriert und die Funktionstüchtigkeit des simulierten Sprechfunks sichergestellt, bevor schließlich die Durchführung des Szenarios starten konnte. Nach dem ersten Szenario wurden die Bedienungen, die für das zweite Szenario relevant waren, trainiert. Anschließend erhielten die Probanden ein Briefing für das zweite Szenario, gefolgt von einer erneuten Kalibrierung des Eye-Trackings. Dann konnte das zweite Szenario in der Simulation erfolgen. Abgeschlossen wurde die Evaluierung in Form einer Nachbesprechung, die auch die Durchführung eines Interviews, der Videokonfrontation und das Ausfüllen subjektiver Fragebögen enthielt.

5.1.5 Bewertung der Bedienung einer Navigationshilfe in einer Machbarkeitsstudie

Im Rahmen der Machbarkeitsstudie ist evaluiert worden, inwiefern die Nutzer von der Einführung einer Navigationshilfe im Vergleich zur bisherigen Cockpit-Auslegung profitieren und ob die entwi-

ckelte Designlösung hierbei sinnvoll eingesetzt werden kann. Bei Bewertung der Gebrauchstauglichkeit wird besonders der Zeitgewinn, der bei Verwendung der Navigationshilfe entsteht, deutlich. Auch die Nutzerzufriedenheit, wie sie mit Hilfe des PSSUQ abgefragt wird, steigt, wenn man die Mediane (vgl. Tabelle 5.3 und Abbildung 5.26 in Kapitel 5.2.5.3) der Machbarkeitsstudie betrachtet. Auch hinsichtlich der Workload zeigt sich, dass sich die Piloten, wenn sie die Navigationshilfe verwenden, weniger belastet fühlen. Zur Verifikation der eingegebenen Wegpunkte nutzen die Piloten jedoch weiterhin die verfügbaren Mittel an Bord in Ergänzung zu der Karte. Nach den Aussagen im Interview dient die Navigationshilfe dem Erhalt des Situationsbewusstseins, da es durch den Zeitgewinn, der durch die Bedienung entsteht, möglich ist, nun schneller den Anweisungen der Luftverkehrskontrolle nachzukommen. Zudem fühlen sich die Piloten unter Nutzung der Navigationshilfe sicherer, die richtigen Navigationspunkte anzufliegen. Zusammenfassend stellte sich in den Interviews heraus, dass die Probanden die Einführung der neuen Navigationshilfe sehr stark befürworten und hinsichtlich des vorgestellten Designs nur kleinere Änderungen anzumerken haben. Insofern sind die identifizierten Änderungen in das Design einzuarbeiten und erneut zu überprüfen.

5.1.6 Bewertung der verwendeten Methoden-Kombination

Mit Bezug auf die verwendeten Methoden des Methoden-Inventars zeigt sich nach Analyse der Ergebnisse, dass der Umfang zur Beschreibung des erwarteten Zugewinns an Nutzerzufriedenheit ausreichend ist. Allerdings stellt sich unter Anwendung der Methoden heraus, dass die Verwendung des monokularen Eye-Tracking-Systems für eine weitere Verwendung fehleranfällig ist. Auch der Aufbereitungs- und Analyseaufwand stellt sich bei Gegenüberstellung des Umfangs der Evaluierung als zu groß dar. Bei Anwendung des NASA TLX wird weiterhin deutlich, dass eine Umkehr der Fragerichtung zur Bewertung der eigenen Leistung als Variable bei den Probanden zu Irritationen führt. Für die weitere Bewertung geeignet erweisen sich stattdessen insbesondere die Erfassung der Nutzerleistung mit Hilfe einer Videokonfrontation, die Erfassung der Eingabezeit sowie der PSSUQ zur Erhebung der Nutzerzufriedenheit. Allerdings zeigt sich die Videokonfrontation als besonders zeitaufwendig. Der PSSUQ stellt sich in der Anwendung als strukturiertes Verfahren heraus, dass auch durch die Möglichkeit an freien Kommentaren die Interviewfragen gut ergänzen kann. Mit den ergänzend verwendeten Interviews ist es möglich, eine Informationstiefe zu erfassen, die mit dem PSSUQ nicht erlangt werden kann, jedoch zeitaufwendig herausgearbeitet werden muss.

5.2 Untersuchung des Re-Designs einer Navigationshilfe in verschiedenen Situationen

Die Gebrauchstauglichkeit einer neu einzuführenden Navigationshilfe wurde in zwei Experimenten evaluiert. In der Machbarkeitsstudie (vgl. Kapitel 5.1) sollte festgestellt werden, ob eine Zeitersparnis durch die Einführung einer Navigationshilfe bewirkt werden kann. Nach Einbindung der Ergebnisse der Machbarkeitsstudie und der Nutzerkommentare in ein Re-Design wurde die Eignung der Funktion für das Cockpit und dessen Gebrauchstauglichkeit erneut in zwei repräsentativen Situationen (Szenarien) mit unterschiedlichem Schwierigkeitsgrad untersucht. Die zentral zu untersuchende Handlung

5 Die Evaluierung des multidimensionalen Methoden-Inventars

des Piloten im Cockpit-Simulator war hier erneut die Anwendung der Navigationshilfe, welche die Eingabe eines Navigationspunktes anhand eines Codes, bestehend aus fünf Buchstaben, ermöglicht.

5.2.1 Stichprobenbeschreibung

Die Navigationshilfe wurde anhand von zehn Piloten mit unterschiedlichem Erfahrungshintergrund, der sowohl Piloten nach gerade abgeschlossener Ausbildung, Testpiloten und operativ tätige Piloten umfasste, evaluiert. Im Mittel entsprach die Erfahrung der Piloten 2327.78 Flugstunden (SD = 1726.29 Stunden). Alle Versuchspersonen erklärten sich mit dem Versuch und der damit verbundenen Weiterverarbeitung ihrer Daten einverstanden und nahmen freiwillig an der Studie teil.

5.2.2 Verwendete Methoden

Zur Evaluierung des Re-Designs einer Navigationshilfe im Cockpit-Simulator wurde ein umfangreiches Methoden-Set, wie in Tabelle 5.2 dargestellt, verwendet.

Tabelle 5.2 Methoden-Inventar zur Erhebung der Qualitätskriterien in der Re-Design-Studie

Methoden	Situationsbewusstsein	Workload	Gebrauchstauglichkeit
	3D SART	NASA TLX	Nutzerleistung
	SAGAT	TDRT	PSSUQ
	Funktionsrelevantes Maß: Sortieraufgabe	Eye-Tracking	Eye-Tracking
	Eye-Tracking		Interviews
			Cognitive Walkthrough

Anmerkung: Die qualitativen Methoden (Interviews und Cognitive Walkthrough) sind nicht Teil der ab Kapitel 6 beschriebenen Validierung

Als strukturierende Methode zur Beschreibung der Funktionsnutzung im Flug wurde eine hierarchische Task-Analyse (HTA) angewendet. Anhand der Eisenhower-Matrix wurden anschließend die Anforderungen an das Situationsbewusstsein kontrastiert und die Szenarien zur Schaffung einer Vergleichsmöglichkeit manipuliert. Zur Variation von Wichtigkeit wurde der Füllstand des Tankes von einem Szenario zum anderen hin verändert. Auch die Dringlichkeit wurde variiert, indem der Pilot in den Szenarien später oder früher Anweisungen erhielt, um die Handlungen im Cockpit durchzuführen. Beide Manipulationsvorgehen sollten den Piloten dazu zwingen, unter höherem (respektive niedrigerem) Zeitdruck die Navigationshilfe zu bedienen. Als weitere vorbereitende methodische Hilfsmittel zur Durchführung der Evaluierung im Cockpit wurden Drehbücher sowie Beobachtungsbögen auf Grundlage der HTA angefertigt.

Zur Ermittlung und Bewertung des Situationsbewusstseins der Piloten während der Evaluierung im Cockpit-Simulator wurde die drei-dimensionale Form der Situation Awareness Rating Technique (3D SART) als subjektiver Fragebogen verwendet sowie retrospektiv angepasste Fragen der Situation Awareness Global Assessment Technique (SAGAT) gestellt und Eye-Tracking als Quelle für objektive Messungen genutzt. Der 3D SART wurde als subjektiver Fragebogen gewählt, da dieser nach den Szenarien schnell anzuwenden, leicht verständlich und als 10D-Anwendung auf einer breiteren Da-

tenbasis in der Luftfahrt (Jones, 2000) zurückzuführen ist. Da der Situationskontext erhalten werden sollte und demnach die verwendeten Methoden möglichst wenig intrusiv einzusetzen waren, wurden die SAGAT-Fragen ebenfalls nach Ende des Szenarios, aber noch im Cockpit, bei ausgeblendeter Karte gestellt. Inhaltlich wurden die Fragen zum Situationsbewusstsein während des Szenarios und der Bedienung der Navigationshilfe anhand einer stummen Karte, direkten Abfragen, Wegpunktlisten und mit dem jeweils zugehörigen Sicherheitsurteil seitens der Probanden erfragt. Um die Handlungsplanung der Piloten zu erfassen, wurde eine Sortieraufgabe als funktionsrelevantes Vorgehen angewendet. Die Sortieraufgabe umfasste die einzelnen Handlungsschritte im Cockpit während der Nutzung der Navigationshilfe. Die einzelnen Handlungsschritte wurden auf Karteikarten gedruckt und den Piloten in durchmischter Reihenfolge dargeboten. Es war Aufgabe des Piloten, nach jedem Szenario die Reihenfolge in den aufgeführten Handlungen zu finden, welche den Tätigkeiten im Cockpit während des Szenarios entsprach. Objektive Messdaten wurden mit Hilfe eines binokularen Eye-Trackers der Firma Ergoneers erhoben und mit der Software D-Lab verarbeitet. Zur Messung des Situationsbewusstseins anhand Blickverhaltens wurden unter anderem Fixationen und deren Dauer, sowohl aufgabenbasiert während der Nutzung der Navigationshilfe sowie generisch während der Gesamtdauer des Szenarios, erhoben.

Um die Workload der Probanden zu bestimmen, wurde der NASA TLX als subjektiver Fragebogen nach jedem Szenario eingesetzt, da dieser zur Evaluierung von Workload in der Luftfahrtbranche bereits eingesetzt und hierfür entwickelt wurde (Hart & Staveland, 1988; Hart, 2006), weiterhin weit verbreitet ist und unkompliziert anzuwenden ist. Zur Operationalisierung des Workloads mit Hilfe von objektiven Messkriterien wurde eine taktile Form der Detection Response Task (TDRT) gewählt sowie erneut das Eye-Tracking verwendet. Die TDRT als Sekundär-Aufgabe wurde taktil angewendet, da davon auszugehen ist, dass der visuelle Kanal des Piloten sehr belastet ist und daher die verbleibende mentale Kapazität über einen anderen Kanal (ISO 17488:2016) wie eben dem taktilen erhoben werden sollte. Umgesetzt wurde die Aufgabe, indem den Piloten eine Vibrationsplatte im Nacken angebracht und ihm ein Druckknopf zur Aufzeichnung seiner Reaktionen und deren Zeiten zur Verfügung gestellt wurde. Die Aufgabe des Piloten war es, während des Szenarios kontinuierlich auf einen zufällig präsentierten taktilen Reiz innerhalb eines Zeitfensters von fünf Sekunden so schnell wie möglich zu reagieren. Auch die Blickdaten des Operators wurden kontinuierlich aufgezeichnet und als objektive Messwerte zur Erhebung von Workload beispielsweise durch die Messung von Pupillendilatation verwendet. Mögliche Ermüdungserscheinungen des Piloten wurden mit Hilfe des PERCLOS (Wierwille et al., 1994) einer Messung des Lidschlussverhaltens erfasst.

Die Erfassung von Gebrauchstauglichkeit wurde mitunter durch die Messung der Eingabezeiten, während die Piloten die Navigationshilfe nutzten, vorgenommen. Zur Erhebung der Nutzerzufriedenheit wurde der PSSUQ eingesetzt. Die Probanden erhielten die Möglichkeit, zusätzlich zu den subjektiven Fragestellungen, freie Kommentare zu äußern, die ebenso erfasst wurden. Der PSSUQ wurde nur einmalig nach Abschluss beider Szenarien angewendet, da eine Neuimplementierung der Navigationshilfe untersucht wurde und somit kein Vergleich zu einem anderen Design möglich war. Auch hinsichtlich der Evaluierung von Gebrauchstauglichkeit wurden Eye-Tracking-Parameter als objektive

5 Die Evaluierung des multidimensionalen Methoden-Inventars

und kontinuierliche Messungen betrachtet. Beispielsweise wurden Transitionen und Sakkaden sowohl aufgabenbasiert als auch generisch dazu genutzt, um die Verteilung an Informationen zu beschreiben und somit das Funktionsdesign zu bewerten. Während des Szenarios konnten nicht alle Aspekte der Nutzung der Navigationshilfe erfasst werden und mussten daher separat erhoben werden. Dies wurde anhand eines sogenannten Cognitive Walkthroughs in Form einer Think-Aloud-Technik umgesetzt. Hierbei sollte der Pilot auch diese Aspekte im Cockpit nutzen und während der Bedienung seine Gedanken frei äußern. Ziel war es, anhand dieser Methode Fehler im Design zu entdecken und später entsprechende Änderungen vorzunehmen. Weiterhin wurden in halbstrukturierten, qualitativen Interviews direkt nach den Szenarios und im Anschluss an die Evaluierung umfassend die Gedanken der Nutzer zur Bedienung der Navigationshilfe und deren Nutzerzufriedenheit erfragt.

5.2.3 Szenario-Beschreibung

Die Cockpit-Evaluation wurde anhand von zwei Szenarien mit unterschiedlicher Schwierigkeit durchgeführt. Den Rahmen beider Szenarien bildete ein Überführungsflug vom Flugplatz Laage zum Flugplatz Manching. In beiden Fällen startete das Szenario an einem festen Navigationspunkt mit einer vordefinierten Flug-Route. Während des Fluges nach Manching erhielt der Pilot die Anweisung, durch die Flugsicherung (ATC) eine veränderte Flug-Route zu fliegen. Zur Eingabe des veränderten Flugweges sollte der Pilot dann die Navigationshilfe nutzen. Kurz vor Erreichen seines Ziellandeplatzes in Manching wurde dem Piloten erneut durch ATC mitgeteilt, dass der Ziellandeplatz und sein Notlandeplatz gesperrt seien und er erneut ausweichen müsse.

Im schwierigeren Szenario wurde der Pilot nach München umgeleitet und musste hier insgesamt sechs Mal die Navigationshilfe nutzen, während er im leichteren Szenario nach Nürnberg umgeleitet wurde und fünf Mal die Navigationshilfe nutzen sollte. In beiden Szenarien erhielt der Pilot durch ATC während des Fluges je vier Anweisungen, die Höhe sowie die Frequenz für Funksprüche zu wechseln.

Wie unter (Kapitel 4.2) erwähnt, wurden die Szenarien auf Grundlage der Eisenhower Matrix hinsichtlich Wichtigkeit und Dringlichkeit manipuliert und somit der zeitliche Druck auf die Piloten bzgl. ihrer Dateneingabe verändert. Demzufolge startete der Pilot im schwierigeren Szenario nach München mit einem geringeren Tankfüllstand (Wichtigkeit) und erhielt Umleitungsanweisungen durch ATC später (Dringlichkeit) als im Nürnberg-Szenario.

5.2.4 Versuchsablauf

Zu Beginn des Versuches wurden die Piloten als Probanden in der Simulationsumgebung willkommen geheißen und erklärten sich nach einer Sicherheitsbelehrung mit der Durchführung des Versuches sowie der Weiterverarbeitung ihrer Daten einverstanden. Die Versuchspersonen erhielten vor Durchführung des Experimentes eine schriftliche Ausarbeitung hinsichtlich der Bedienung der Navigationshilfe und wurden hierzu am Tag der Untersuchung mit Hilfe einer Präsentation im De-/Briefing-Raum zusätzlich eingewiesen. Nachdem erste Rückfragen seitens des Piloten geklärt wurden, hatte dieser die Möglichkeit, die Bedienung der Navigationshilfe im Cockpit-Simulator umfassend zu trainieren.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Nach Abschluss des Trainings wurde der Pilot, so weit möglich, anhand einer Präsentation über die An- und Verwendung der unter Kapitel 5.2.2 aufgeführten Methoden aufgeklärt. Die Vorbereitung auf das im Simulator zu fliegende Szenario fand anschließend mit Hilfe einer weiteren Präsentation statt, die den Piloten mit den hierfür notwendigen Situationsverständnis ausstatten sollte. Das heißt, der Pilot erhielt beispielsweise Informationen zu Start- und Zielflugplatz, dem Füllgrad des Tankes sowie Wetterinformationen. Im Cockpit-Simulator wurde dann zunächst eine Baseline bezüglich der TDRT erhoben und der Eye-Tracker kalibriert. Anschließend wurde das erste Szenario, je nach Randomisierung entweder nach München oder nach Nürnberg, geflogen. Währenddessen wurde nicht nur das Flugverhalten der Probanden kontinuierlich aufgezeichnet, sondern auch deren Eingabezeiten, ihr Blickverhalten und ihre Reaktionen auf die TDRT. Nach Beendigung des Szenarios wurde jeder Pilot zu seinen situativen Belastungen und seinem Zugang zu Informationen in einem qualitativen Situationsinterview befragt. Anschließend erhielten die Probanden in durchmischter Reihenfolge eine Sortieraufgabe, den 3D SART, angepasste SAGAT-Fragen und den NASA TLX. In diesem Zeitraum hatte er die Möglichkeit, die Eye-Tracking-Brille abzulegen und nach dem Ausfüllen der Fragebögen zu pausieren. Ehe das nächste Szenario dann im Anschluss geflogen werden konnte, wurde der Eye-Tracker erneut kalibriert. Schließlich wurde nach Beendigung des zweiten Szenarios das methodische Vorgehen wiederholt. Dann wurden weitere Aspekte der Funktion, die nicht im Szenario integriert waren, anhand eines Cognitive Walkthroughs evaluiert. In einer abschließenden Nachbesprechung sollte der Proband anhand von einer Videokonfrontation einzelne Ereignisse des Szenarios nachvollziehen und auf dieser Grundlage die Funktion hinsichtlich deren Gebrauchstauglichkeit mit Hilfe des PSSUQ und weiteren Interviewfragen bewerten. Dies markierte den Schlusspunkt der Evaluierung und der Proband wurde entlassen.

5.2.5 Bewertung der Verwendung des Re-Designs einer Navigationshilfe

Das Re-Design der Navigationshilfe ist in der Auswertung unter zwei verschiedenen Aspekten betrachtet worden:

- Schwierigkeit des Szenarios
- Leistungsverbesserung durch Erlernen der Bedienung der Navigationshilfe

Ursprünglich wurde die Bedienung der Navigationshilfe unter Berücksichtigung der Szenario-Schwierigkeit untersucht. Nach Analyse der deskriptiven Ergebnisse stellte sich heraus, dass der Einfluss der Szenario-Schwierigkeit auf die Bedienung der Navigationshilfe und die Qualitätskriterien ohne Bedeutung ist – die Navigationshilfe also bei unterschiedlicher Schwierigkeit gleich gut genutzt werden kann.

Die Daten wurden daher in einem weiteren Schritt unter dem Aspekt der verbesserten Bedienung im Vergleich vom ersten zum zweiten Szenario interpretiert. Da im Folgenden beispielhaft erläutert werden soll, in welcher Form die Daten in den Validierungsprozess eingegangen sind, wird nur eine Interpretation der Daten ausführlich beschrieben – die Leistungsverbesserung durch Erlernen der Bedienung der Navigationshilfe. Aufgrund der Stichprobengröße und der unterschiedlichen Anzahl an Messpunkten wird auf eine rein deskriptive Berichterstattung zurückgegriffen.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

5.2.5.1 Bewertung von Situationsbewusstsein im Rahmen der Evaluierung eines Re-Designs

Um die Bedeutung des Re-Designs für das Situationsbewusstsein der Nutzer zu erfassen, sind subjektive und objektive Messdaten erhoben worden.

3D Situation Awareness Rating Technique (3D SART)

Ein subjektiver Fragebogen ist der 3D SART, der Situationsbewusstsein anhand der Kriterien ‚Bedarf an Aufmerksamkeits-Ressourcen‘, ‚Verfügbarkeit von Aufmerksamkeits-Ressourcen‘ und ‚Verständnis der Situation‘ operationalisiert. Aus diesen drei erfragten Werten mit einer Abstufung von 1 (niedrig) bis 7 (hoch) kann ein Gesamtwert zur Bestimmung des Situationsbewusstseins mit der Formel (Verständnis – (Bedarf – Verfügbarkeit)) errechnet werden, der den Wertebereich -5 bis +13 annehmen kann. In der Untersuchung des Re-Designs einer Navigationshilfe lässt die Erhebung darauf schließen, dass sich das Erlernen der Bedienung der Navigationshilfe auf das empfundene Situationsbewusstsein der Piloten niederschlägt, da diese im zweiten Szenario über eine geringere Beanspruchung (Mdn = 4.5) und dafür über eine höhere Verfügbarkeit von Aufmerksamkeits-Ressourcen (Mdn = 6) berichten, als im Vergleich zum ersten Szenario (Mdn Bedarf an Aufmerksamkeits-Ressourcen = 5.5, Mdn Verfügbarkeit an Aufmerksamkeits-Ressourcen = 4.5). In beiden Szenarien bewerten die Piloten ihr Situationsverständnis als hoch, wobei im zweiten Szenario das Urteil nahezu einheitlich mit dem Maximalwert 7 bewertet wird und im ersten Szenario die Probanden zwischen den Werten 6 und 7 schwanken. Abbildung 5.1, welche die deskriptive Datenverteilung des ersten Szenarios zeigt, verdeutlicht im Vergleich zu den Boxplots des zweiten Szenarios in Abbildung 5.2 weiterhin, dass die Nutzer im zweiten Szenario einheitlicher davon überzeugt sind, durch die Navigationshilfe in ihren Aufmerksamkeits-Ressourcen gestützt zu werden.

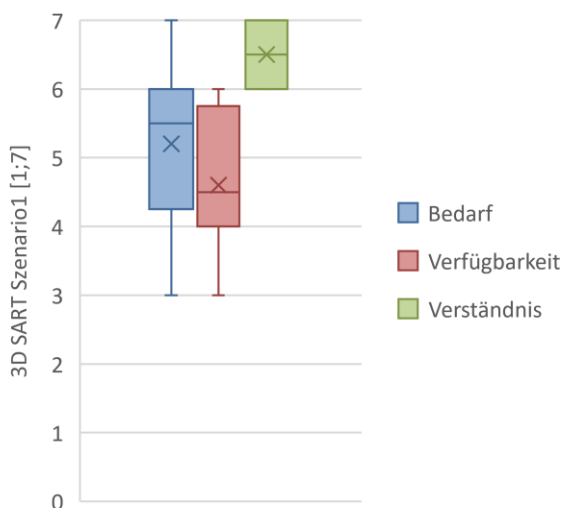


Abbildung 5.1 Boxplots des 3D SART nach dem 1. Szenario

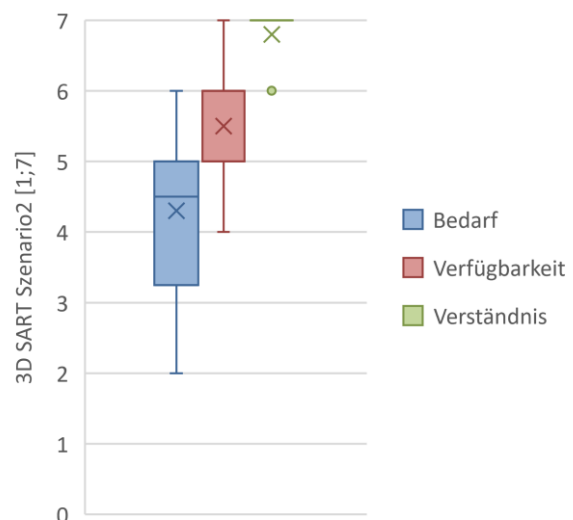


Abbildung 5.2 Boxplots des 3D SART nach dem 2. Szenario

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Dieser Eindruck spiegelt sich folglich auch bei Betrachtung des berechneten Gesamtwertes des Situationsbewusstseins in Abbildung 5.3 wider. Dementsprechend wird das Situationsbewusstsein von den Piloten im zweiten Szenario (grüne Boxplots) höher eingestuft (Minimum = 5, Mdn = 8, Maximum = 10) als im ersten Szenario (Minimum = 4, Mdn = 5.9, Maximum = 9).

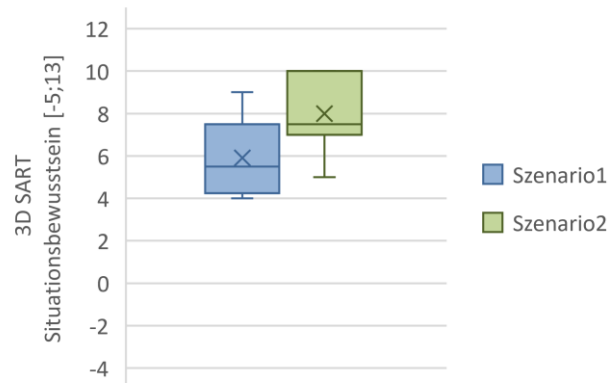


Abbildung 5.3 Boxplots der 3D SART-Gesamtwerts (Situationsbewusstsein) in Abhängigkeit der Szenario-Reihenfolge

Situation Awareness Global Assessment Technique (SAGAT)

Ergänzend zu den Aussagen des 3D SARTs ist das Situationsbewusstsein anhand von angepassten Fragen des SAGAT mittels stummer Karten sowie Fragen zu den Navigationspunkten und den Ereignissen im Szenario ermittelt worden.

Das Reizmaterial der Stummen Karte wird beispielhaft in Abbildung 5.4 (links) für das Münchner Szenario dargestellt. Die abgebildeten roten Rechtecke sind hierbei als Fixpunkte zu verstehen, die den Piloten zur Orientierung dienen sollten. Als Fixpunkte sind in beiden Szenarien die Startpunkte des jeweiligen Szenarios angegeben worden, der anvisierte Ziellandeplatz sowie der Ausweichlandeplatz nach Bekanntgabe der Umleitung. Die blauen Punkte stellen hingegen 30 potenzielle Lösungspunkte bzw. Distraktoren dar. In der rechten Abbildung 5.5 (Münchner Szenario) ist mittels der hellgrünen Punkte die Musterlösung dargestellt sowie der Lösungsvorschlag von Versuchsperson (VP) 8 (hellblaue Linie).

5 Die Evaluierung des multidimensionalen Methoden-Inventars

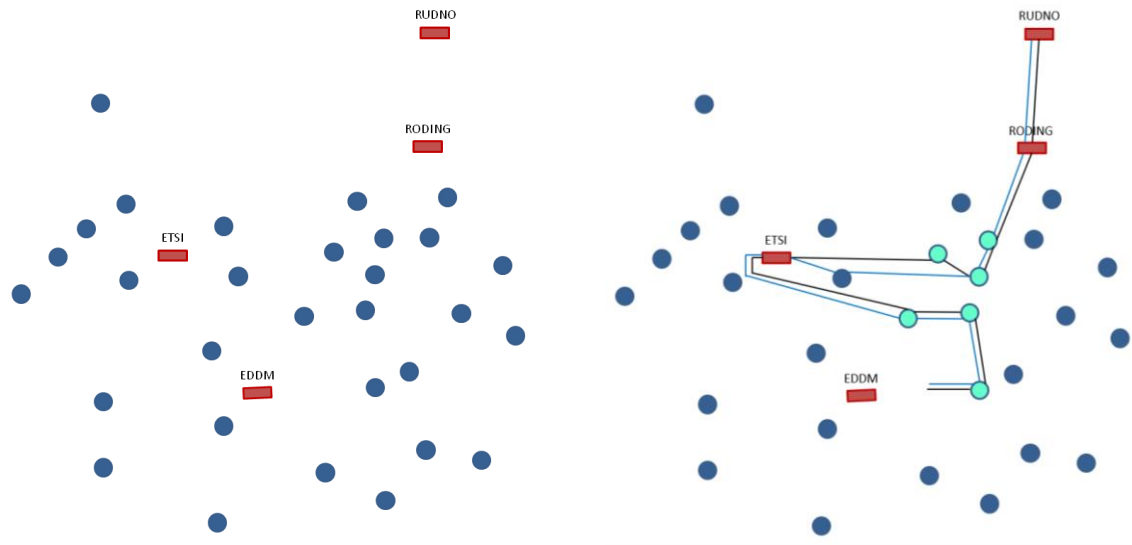


Abbildung 5.4 SAGAT-Reizmaterial der stummen Karte mit Fixpunkten für das Szenario München

Abbildung 5.5 SAGAT-Lösung der stummen Karte für das Szenario München

Anmerkung: SAGAT-Lösung (hellgrüne Punkte und schwarze Linie) mit Lösungsvorschlag von VP 8 für das Szenario München (hellblaue Linie)

Aufgabe der Piloten war es nicht nur, die Lage der Punkte zu identifizieren und die korrekte Route einzuzeichnen, sondern sich auch an die korrekte namentliche Bezeichnung der Navigationspunkte zu erinnern und diese zuzuweisen. Hieraus ergeben sich für die Auswertung drei Aspekte, die mit diesen SAGAT-Fragen betrachtet werden:

- Korrekte Routenabschnitte
- Korrekt erinnerte Navigationspunkte ohne Namensbezug / Buchstabencode
- Korrekt erinnerte Position der Navigationspunkte mit Namenszuweisung / Buchstabencode

Für den in Abbildung 5.5 gezeigten Lösungsvorschlag bedeutet das, dass VP 8 fünf von sieben Routenabschnitten wiedergegeben hat und damit zu 71 % der Musterlösung entspricht. Bezogen auf die erinnerten Navigationspunkte ohne Namensbezug ergibt sich, dass der Proband fünf von sechs Navigationspunkten richtig, d.h. mit 83 % identifizieren konnte. Da VP 8 nur in der Hälfte aller Fälle den korrekten Namen des Wegpunktes der Position auf der stummen Karte zugeordnet hat, ergibt sich hier eine SAGAT Leistung von 50 %.

Die SAGAT-Gesamtleistung der Probanden bezüglich der Szenario-Reihenfolge wird in Abbildung 5.6 dargestellt. Deutlich wird, dass die Probanden im zweiten Szenario (grüne Boxplots) bzgl. der drei Auswertevorgänge jeweils zu einer besseren Orientierungs- und Erinnerungsleistung tendieren als im Vergleich zum ersten Szenario (blaue Boxplots). Es zeigt sich des Weiteren eine Verringerung der Spannweite bei Auswertung der Routenabschnitte sowie insbesondere bei den erinnerten Navigationspunkten ohne Buchstabencode. Weiterhin ist zu sehen, dass die korrekte Identifizierung der Position der Navigationspunkte mit Namenszuweisung (Boxplots rechts) am schwersten zu bewältigen war, da hier die Leistung der Probanden geringer ist (Mdn Szenario 1 = 35 %; Mdn Szenario 2 = 60 %).

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Eine Verbesserung der SAGAT Leistung ist, sofern die Szenario-Reihenfolge betrachtet wird, sowohl anhand der verbesserten Nutzung der Navigationshilfe als auch mit dem Erlernen der Methode zu erklären.

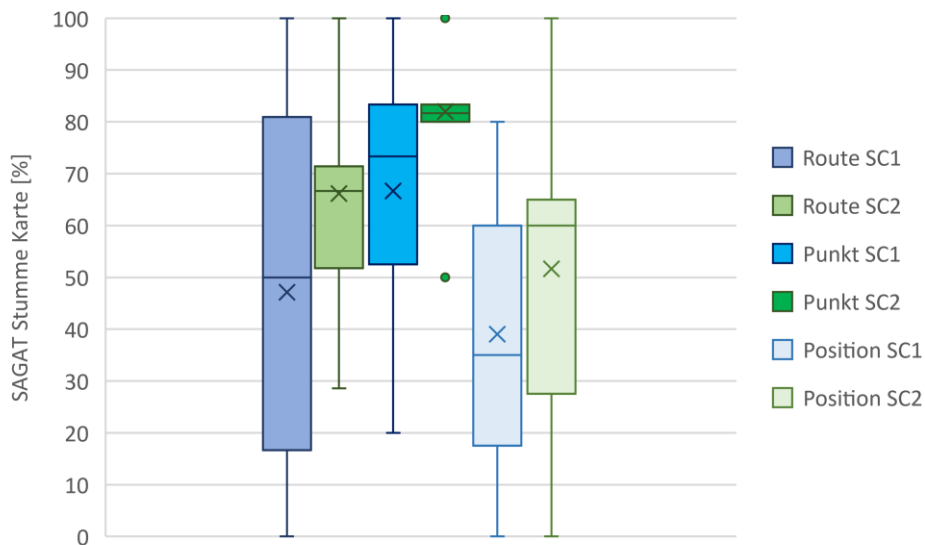


Abbildung 5.6 Boxplots der SAGAT-Antwortleistung in Abhängigkeit der Szenario-Reihenfolge

Anmerkung: Die Bezeichnung Route steht hierbei für die richtig erkannten Routenabschnitte im ersten Szenario (linker blauer Boxplot) im Vergleich zum zweiten Szenario (linker grüner Boxplot). Die beiden mittleren Boxplots zeigen den Vergleich bezüglich der richtigen Identifikation der Navigationspunkte ohne Namenszuweisung zwischen den beiden Szenarien. Ganz rechts wird die Antwortleistung der Probanden hinsichtlich der richtigen Positionierung der Navigationspunkte auf der stummen Karte mit Namenszuweisung dargestellt.

Neben der Verortung der Navigationspunkte auf der stummen Karte bewerteten die Piloten auch, wie sicher sie sich hinsichtlich dieses Urteils fühlten (0 = weiß ich nicht; 1 = sehr unsicher; 5 = sehr sicher).

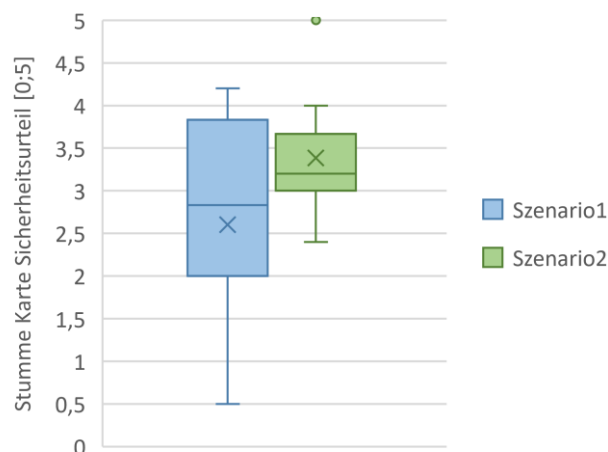


Abbildung 5.7 Boxplots des SAGAT-Sicherheitsurteils in Abhängigkeit der Szenario-Reihenfolge

Unter Betrachtung von Abbildung 5.7 ist zu erkennen, dass sich das Sicherheitsurteil der Probanden im Vergleich zum ersten Szenario (Mdn = 2.83) im zweiten Szenario (Mdn = 3.20) minimal ver-

5 Die Evaluierung des multidimensionalen Methoden-Inventars

bessert hat. Deutlich hingegen scheint, dass sich das Urteilsverhalten der Probanden im zweiten Szenario angleicht und einheitlicher wird. Eine mögliche Begründung hierfür könnte eine verbesserte Kompetenz in der Bedienung der Navigationshilfe sein, ist aber auch mit der Möglichkeit zu erklären, dass die Piloten im zweiten Szenario mehr Aufmerksamkeit auf diese Fragestellung lenkten.

Neben der stummen Karte wurden den Probanden auch Fragen zum Szenario-Verlauf gestellt, die Teil der jeweiligen Navigationsaufgaben waren – also inhaltlich auf Frequenzwechsel, Luftdruck- sowie Höhenänderungen abzielten und ebenso unter der Methode SAGAT zusammengefasst wurden.

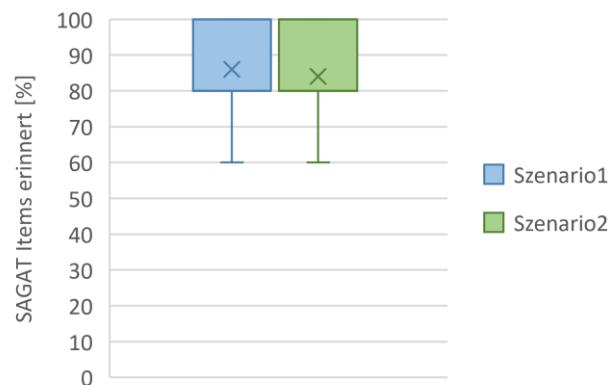


Abbildung 5.8 Boxplots der SAGAT-Items in Abhängigkeit der Szenario-Reihenfolge

Es zeigt sich, wie in Abbildung 5.8 dargestellt, dass sich die verwendeten SAGAT-Items nicht zur Diskriminierung der Szenarien-Reihenfolge eignen, da die Probandenleistung in diesem Fall nahezu identisch und vor allem übermäßig gut ist. Hieraus lässt sich zum einen ableiten, dass die gewählten Fragestellungen zu einfach waren und weiterhin schlussfolgern, dass zu wenige Items verwendet wurden, um für eine höhere Datenvariabilität zu sorgen.

Ergänzend zu den genannten Verfahren zur Bestimmung der SAGAT Leistung sind den Probanden Listen mit 30 Buchstabencodes verschiedener Navigationspunkte ausgehändigt worden. Aufgabe der Piloten war es hier, die Buchstabencodes zu identifizieren, die während der Szenarien relevant waren und diejenigen Navigationspunkte zurückzuweisen, die nicht Teil des Szenarios waren. Wie in Abbildung 5.9 zu erkennen ist, eignet sich dieses Vorgehen nicht, um den Lernfortschritt von einem Szenario zum anderen zu skizzieren. Sowohl die Identifizierung der genannten Buchstabencodes als auch die richtige Zurückweisung der nicht relevanten Buchstabencodes findet auf einem hohen Niveau statt und ist hinsichtlich der Szenarien-Reihenfolge nahezu identisch.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

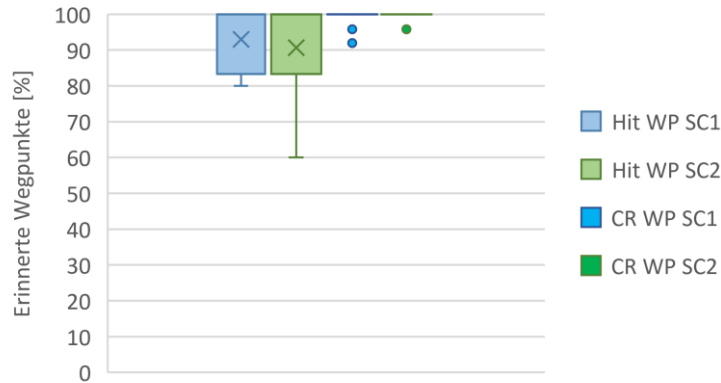


Abbildung 5.9 Boxplots der mittels SAGAT erinnerten Navigationspunkte in Abhängigkeit der Szenario-Reihenfolge

Anmerkung: Darstellung des prozentualen Anteils korrekt erinnerte Navigationspunkte (grüner und blauer Boxplot links) und korrekt zurückgewiesene Navigationspunkte (dunkelblauer und dunkelgrüner Boxplot rechts)

Auch in diesem Fall sind die Piloten zusätzlich dazu befragt worden, wie sicher sie sich in ihrem Urteil sind. Es stellt sich heraus (vgl. Abbildung 5.10), dass sich die Piloten nur bezüglich der richtig identifizierten Wegpunkte im zweiten Szenario (grüner linker Boxplot) prozentual gesehen geringfügig unsicherer (Mdn = 88 %) sind als im ersten Szenario (Mdn = 100 %). Im Hinblick auf das Sicherheitsurteil der korrekt zurückzuweisenden Navigationspunkte zeigt sich hingegen kein augenscheinlicher Unterschied.

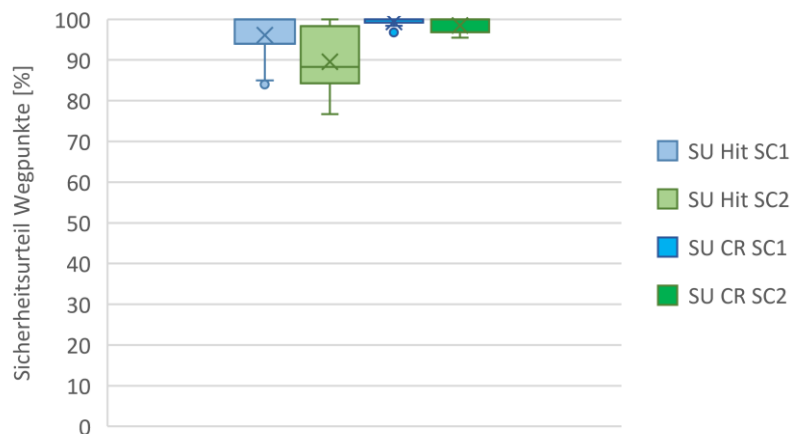


Abbildung 5.10 Boxplots der SAGAT-Sicherheitsurteile bezüglich der Navigationspunkte in Abhängigkeit der Szenario-Reihenfolge

Anmerkung: Darstellung des prozentual gewürdigten Sicherheitsurteils korrekt erinnerte Navigationspunkte (grüner und blauer Boxplot links) und korrekt zurückgewiesene Navigationspunkte (dunkelblauer und dunkelgrüner Boxplot rechts)

Die Ergebnisse dieses Vorgehens im Rahmen des SAGATs lassen wiederum den Schluss zu, dass diese Befragung zur korrekten Identifikation oder Zurückweisung der Navigationspunkte zu leicht war, um hinsichtlich des Erlernens der Navigationshilfe von einem Szenario zum anderen zu unterscheiden.

Sortieraufgabe

Zur Überprüfung der stichprobenabhängigen Handlungsplanung und -vorgehens ist eine Sortieraufgabe eingeführt worden. Hierzu sind die Cockpit-Tätigkeiten, die während der Bedienung der Navigationshilfe zusätzlich durchgeführt werden, einzeln auf Karteikarten geschrieben worden und dem Probanden zur Sortierung vorgelegt worden. Jeweils in Abhängigkeit der Szenarien bzw. der Szenarien-Reihenfolge ist anhand der Stichprobenergebnisse eine optimale Handlungsabfolge abgeleitet worden, die auf Grundlage der Platzierungshäufigkeit der Tätigkeit zu ermitteln war. Dieses Prinzip ist auch zur Ermittlung einer idealen Handlungsabfolge übernommen worden, in welcher alle Stichprobenantworten, unabhängig von den Szenarien, berücksichtigt werden konnten. Als Probandenleistung ist dann die jeweilige relative Passung der Einzelergebnisse zu der konsensbasierten ‚optimalen‘ Handlungsabfolgen zu verstehen – zum einen in Anbetracht der Szenario-Reihenfolge und zum anderen im Hinblick auf eine ideale Handlungsabfolge.

Wie in Abbildung 5.11 gezeigt, lässt die Szenario-Reihenfolge keine weiteren Rückschlüsse zu (Mdn Szenario 1 = 0.39, Mdn Szenario 2 = 0.36), außer, dass die Probanden im zweiten Szenario bezüglich ihrer Handlungsreihenfolge zu marginal einheitlicheren Ergebnissen gelangt sind.

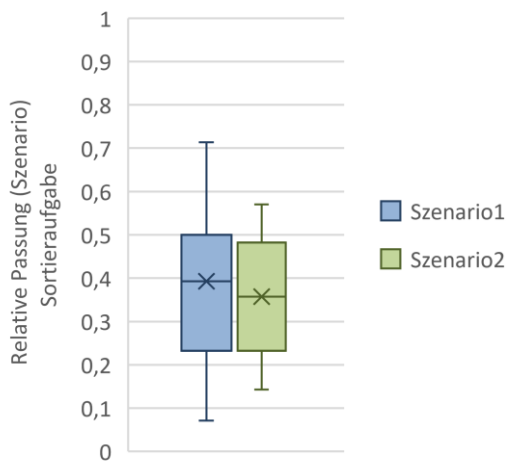


Abbildung 5.11 Boxplots der relativen Passung der Probandenantworten in der Sortieraufgabe in Abhängigkeit der Szenario-Reihenfolge

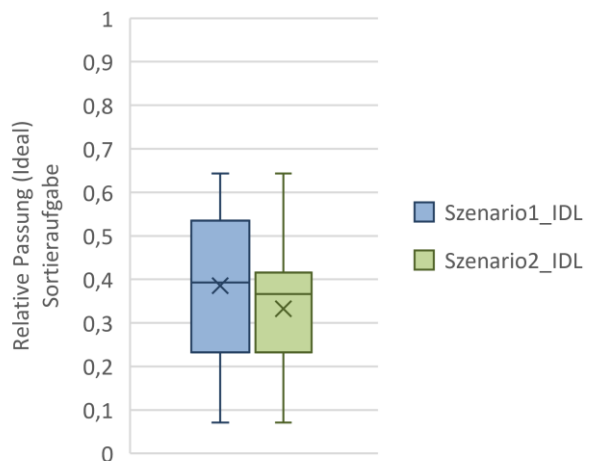


Abbildung 5.12 Boxplots der relativen Passung der Probandenantworten zur Ideallösung in der Sortieraufgabe in Abhängigkeit der Szenario-Reihenfolge

Betrachtet man darüber hinaus die relative Passung im Hinblick auf die ideale Handlungsabfolge, zeigt sich, dass diese im zweiten Szenario unwesentlich geringer ausfällt (Mdn = 0.37) als im ersten Szenario (Mdn = 0.39), da die Hälfte der Daten unterhalb des Medians stark streuen, das dritte Quartil aber einen eher engen Bereich umschreibt und die Range der Boxplots nahezu identisch bleibt (Abbildung 5.12). Inhaltlich gesehen wäre hier zu vermuten, dass die Probanden sich auf Grund der verbesserten Bedienung der Navigationshilfe nun vermehrt auf ihre individuellen Flugverfahren und ihre Erfahrung verlassen, als im Vergleich zum ersten Szenario. Eine weitere Betrachtung dieser Methode wäre daher durchaus gerechtfertigt.

Eye-Tracking

Als objektive Komponente zur Operationalisierung von Situationsbewusstsein ist das Blickverhalten der Probanden gemessen worden, da unter Berücksichtigung der Eye-Mind-Hypothese (Just & Carpenter, 1980) explizit wird, wie die Informationsaufnahme durch den Probanden erfolgt und wie stark dessen Aufmerksamkeit gebunden ist. Fixationen sind deshalb als Indikator für Situationsbewusstsein geeignet, da sie durch Auftreten und Dauer die Informationsaufnahme widerspiegeln (s.a. Moore, 2009). Unterschieden wird im vorliegenden Vorgehen zwischen der aufgabenbasierten Informationsaufnahme (Abbildung 5.13 und Abbildung 5.15), in welcher nur das Blickverhalten während der Nutzung der Navigationshilfe betrachtet wird und der Erfassung des Blickverhaltens während der Szenarien in ihrer Gänze. In Abhängigkeit der Szenarien-Reihenfolge ist in Abbildung 5.13 und Abbildung 5.14 zu erkennen, dass die Anzahl an Fixationen (Abbildung 5.13) im zweiten Szenario geringer ist (Mdn = 566.5) als im ersten Szenario (Mdn = 854) und die Dauer der Fixationen (Abbildung 5.14) im zweiten Szenario vergleichsweise ansteigt (Mdn Szenario 1 = 249.03 ms, Mdn Szenario 2 = 334.9 ms). Dies lässt die Vermutung zu, dass die Informationsaufnahme im zweiten Szenario von den Probanden effizienter umgesetzt werden kann und ermöglicht den Rückschluss auf eine verbesserte Nutzung der Navigationshilfe.

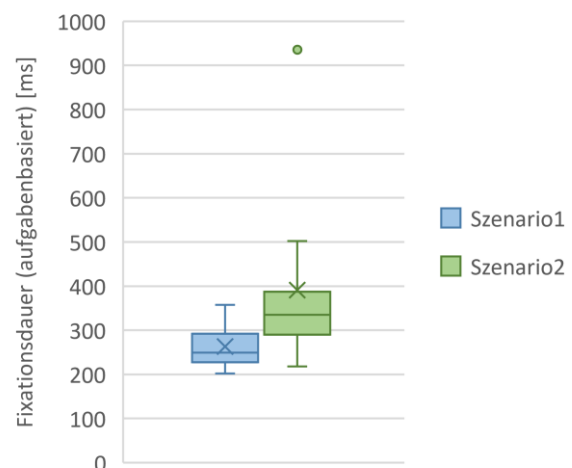
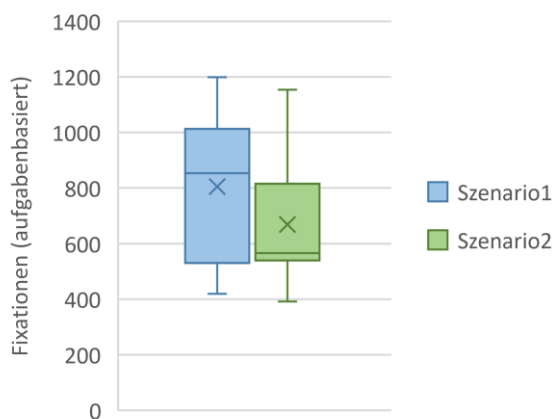


Abbildung 5.13 Anzahl der Fixationen während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge

Abbildung 5.14 Fixationsdauer während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge

Das Gesamtblickverhalten (Abbildung 5.15 und Abbildung 5.16), welches innerhalb eines ganzen Szenarios erfasst worden ist, unterscheidet sich insofern, dass die Anzahl an Fixationen (Abbildung 5.15) im zweiten Szenario variabler wird. Das heißt, obwohl sich die Mediane vom ersten Szenario (Mdn = 3177) zum zweiten Szenario (Mdn = 3160.5) kaum unterscheiden, streuen die Daten doch stärker, jedoch in beide Richtungen. Es wiederholt sich hier das Phänomen, dass im zweiten Szenario die Dauer der Fixationen (Mdn = 329.19 ms) erneut im Vergleich zum ersten Szenario (Mdn = 266.04 ms) zugenommen hat (Abbildung 5.16). Die Ergebnisse sprechen daher dafür, dass sich die Intensität der Informationsaufnahme verändert hat, die Strategie zur Bearbeitung des Szenarios jedoch nicht.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

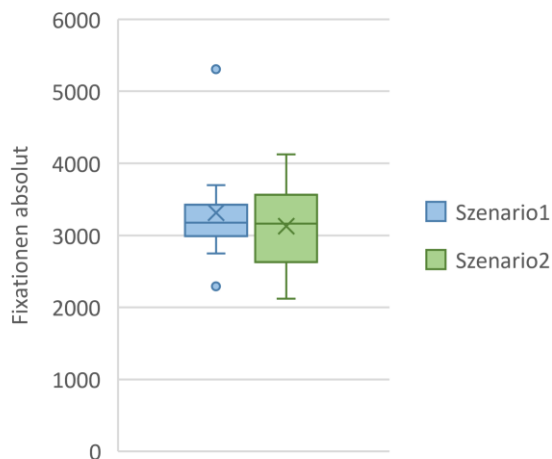


Abbildung 5.15 Anzahl der Fixationen während der Szenarien in Abhängigkeit der Szenario-Reihenfolge

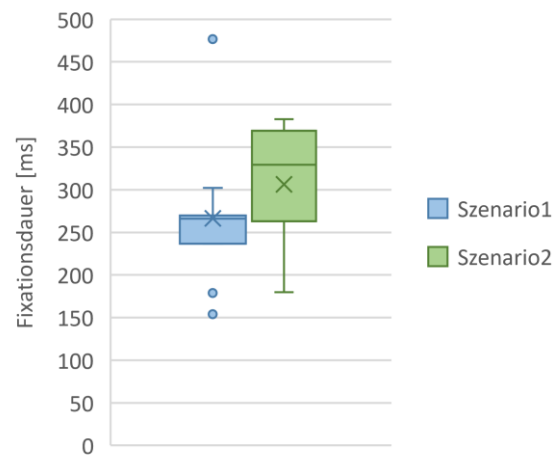


Abbildung 5.16 Fixationsdauer während der Szenarien in Abhängigkeit der Szenario-Reihenfolge

Zusammenfassung der Bewertung des Situationsbewusstseins

In der Gesamtschau auf das Situationsbewusstsein ist zu erkennen, dass das Erlernen der Bedienung der Navigationshilfe von einem Szenario zum zweiten Szenario bei den Probanden dazu führt, ihr Situationsbewusstsein als besser einzuschätzen. Diese Erkenntnis stützen sowohl die subjektiven Daten, die beispielsweise durch den 3D SART erhoben worden sind, wie auch die objektiveren Leistungen des SAGATs bzgl. der Stummen Karte sowie die Erfassung des Blickverhaltens. Wenig zu dieser Erkenntnis beigetragen haben hierbei die Methode der Sortieraufgabe sowie die Befragung zu den SAGAT Items und den Navigationspunkten unter Einbindung von Distraktoren. Eine kontextabhängige Überarbeitung dieser Vorgehen in Abhängigkeit der zu überprüfenden Cockpit-Modifikation ist daher für weitere Evaluierungsvorgehen zur Erhebung von Situationsbewusstsein zu empfehlen.

5.2.5.2 Bewertung der Workload im Rahmen der Evaluierung eines Re-Designs

Die Operationalisierung von Workload zur Evaluierung eines Re-Designs wird, ebenso wie das Situationsbewusstsein und die Gebrauchstauglichkeit mit subjektiven und objektiven Instrumenten vorgenommen.

NASA TLX

Der NASA TLX ist im vorliegenden Fall als subjektiver Fragebogen verwendet worden und mit einer Skala von 0 (gering) bis 10 (sehr stark) in .5er-Schritten erfasst worden. Eine Gewichtung der einzelnen Dimensionen ist nicht vorgenommen worden, da die Gewichtung die Ergebnisse möglicherweise verzerrt (Nygren, 1991) und dies weiterhin aus versuchsökonomischen Gründen nicht gerechtfertigt schien. Dennoch kann ein Workload-Gesamtwert gebildet werden, der als Mittelwert über alle Workload-Dimensionen hinweg für jede Versuchsperson errechnet worden ist. Legt man die Szenarien-

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Reihenfolge als Unterscheidungsmerkmal zu Grunde, werden unter Betrachtung von Abbildung 5.17 für das erste Szenario und Abbildung 5.18 für das zweite Szenario folgende Unterschiede ersichtlich:

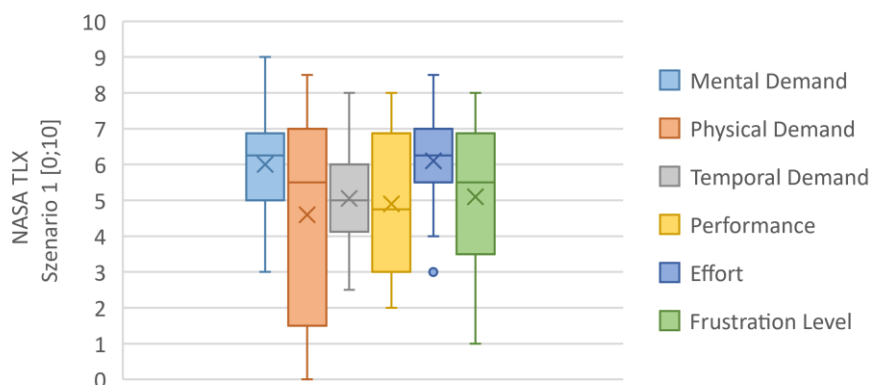


Abbildung 5.17 NASA TLX Boxplots der Workload-Dimensionen des ersten Szenarios

- **Mentale Beanspruchung (hellblaue Boxplots):** Werden nur die Mediane der mentalen Beanspruchung verglichen, ist von einem leichten Unterschied zwischen dem ersten und dem zweiten Szenario auszugehen (Mdn Szenario 1 = 6.25, Mdn Szenario 2 = 4.25). Auch, dass die Extremwerte im zweiten Szenario jeweils geringer ausfallen als im ersten Szenario deutet darauf hin. Jedoch zeigt sich, dass die Antworten im zweiten Szenario stärker streuen und sich somit das Erlernen der Bedienung in der Stichprobe unterschiedlich stark niederschlägt. Das könnte darin begründet sein, dass Piloten mit unterschiedlichem Erfahrungsstand verschieden stark vom Erlernen der Navigationshilfe profitieren.
- **Physische Beanspruchung (orange Boxplots):** Bezüglich der physischen Beanspruchung der Piloten ist nur ein leichter Nachteil bei der Bedienung im ersten Szenario (Mdn = 5.5) im Vergleich zum zweiten Szenario (Mdn = 4.75) zu erkennen. Augenscheinlich hingegen ist, dass die Aussagen zur physischen Beanspruchung für das erste Szenario sehr stark variieren und die Probanden im zweiten Szenario etwas einheitlicher mit einer Tendenz zu geringeren Workload-Werten antworten. D.h. ein Erlernen der Bedienung schlägt sich zwar in den Werten wieder, wird jedoch von den Probanden unterschiedlich stark wahrgenommen.
- **Zeitliche Beanspruchung (graue Boxplots):** Bei Betrachtung der zeitlichen Beanspruchung der Piloten während der Bedienung der Navigationshilfe ist kaum ein Unterschied zwischen den beiden Szenarien ersichtlich. Lediglich eine leichte Tendenz zu höheren Workload-Werten ist für diese Dimension für das erste Szenario zu erkennen (Mdn Szenario 1 = 5, Mdn Szenario 2 = 4.25)
- **Leistung (gelbe Boxplots):** Die Piloten bewerten ihre Eigenleistung im zweiten Szenario (Mdn = 2) deutlich besser als im ersten Szenario (Mdn = 4.75) und zeigen in diesem Urteil auch deutlich mehr Konsens.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

- *Ressourcen-Einsatz (blaue Boxplots)*: Der Aufwand, der zu erbringen war, um die jeweilige Leistung zu erzielen, ist für das zweite Szenario deutlich variabler eingeschätzt worden als für das erste Szenario. Dies spiegelt sich auch in einem leicht geringeren Median für das zweite Szenario (Mdn = 5) als für das erste Szenario (Mdn = 6.25) wider. Die angegebenen Extremwerte liegen jedoch im zweiten Szenario immer unterhalb der Werte (Minimum = 1.5, Maximum = 8) wie sie im ersten Szenario (Minimum = 3, Maximum = 8.5) vorliegen. Ein möglicher Ansatz, dieses Ergebnis zu erklären, liegt darin, dass der Grad an Erfahrung innerhalb der Stichprobe sehr unterschiedlich ist und dementsprechend der Ressourcen-Einsatz bei erfahreneren Piloten im zweiten Szenario kleiner ist als beispielsweise bei weniger erfahreneren Probanden.
- *Frustrationslevel (grüne Boxplots)*: Auch die empfundene Frustration bei der Bedienung der Navigationshilfe scheint im zweiten Szenario geringer (Mdn = 2.5) zu sein als im ersten (Mdn = 5.5). Jedoch zeigen die Boxplots (rechte grüne Boxplots) der Szenarien in beiden Fällen eine deutliche Streuung der Daten an.

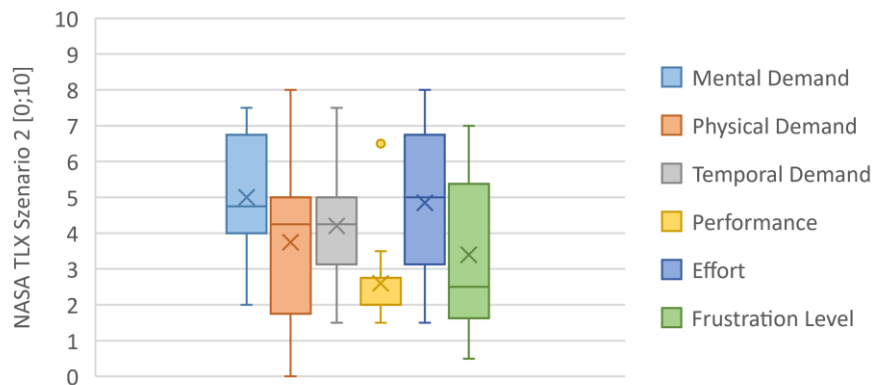


Abbildung 5.18 NASA TLX Boxplots der Workload-Dimensionen des zweiten Szenarios

Auch wenn es in den Einzeldimensionen nicht immer offensichtlich ist, ist zu erkennen, dass die empfundene Workload vom ersten zum zweiten Szenario sinkt; d.h. dass das Erlernen der Bedienung der Navigationshilfe zu einer geringeren empfundenen Workload geführt haben könnte. Dieser Eindruck wird zwangsläufig verstärkt, wenn man den ermittelten Workload-Gesamtwert in Abhängigkeit der Szenarien-Reihenfolge in Abbildung 5.19 betrachtet. Auch hier wird deutlich, dass die Probanden dazu neigen, für das erste Szenario höhere Workload-Werte (Mdn = 5.12) anzunehmen als für das zweite Szenario (Mdn = 3.63). Allerdings fallen im ersten Szenario vier Werte als Ausreißer ins Gewicht. Der Boxplot für das erste Szenario würde allerdings auch dann eine Tendenz zu einer höheren Workload indizieren, wenn diese Werte nicht als Ausreißer behandelt worden wären. Zumindest würde das Minimum von 3.33 des Boxplots des ersten Szenarios dem 0.25 Quartil des zweiten Szenarios von 3.06 nahezu entsprechen.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

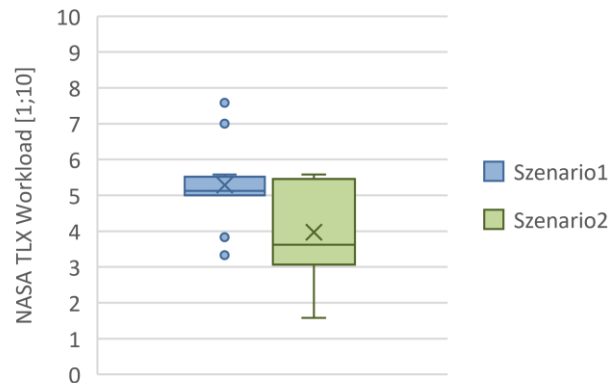


Abbildung 5.19 Boxplots des NASA TLX Gesamt-Workloads in Abhängigkeit der Szenario-Reihenfolge

Tactile Detection Response Task (TDRT)

Als objektive Methodik zur Erhebung der Workload während der Bedienung einer re-designten Navigationshilfe wurde eine taktile Form der DRT gewählt, welche dem Zweitaufgaben-Paradigma entspringt. Der taktile Reiz ist anhand einer Vibrationsplatte, die im Nacken des Probanden befestigt worden ist, appliziert worden. Die zu erfassenden Daten basieren auf den Reaktionszeiten innerhalb eines 5-Sekunden-Zeitfensters und den entsprechenden Trefferquoten. Die Werte sind jeweils während der beiden aufeinanderfolgenden Szenarien und in einer Baseline-Messung erfasst worden. Offensichtlich ist, dass bei Betrachtung der Reaktionszeiten (Abbildung 5.20) und der relativen Hit-Rate (Abbildung 5.21) die Probanden deutlich davon profitieren, dass keine Primäraufgabe vorhanden ist. Zwischen den beiden Szenarien kann hingegen weder bei den Reaktionszeiten (Mdn Szenario 1 = 1641.67 ms, Mdn Szenario 2 = 1720.8 ms) noch bei den Hit-Rates (Mdn Szenario 1 = 68 %, Mdn Szenario 2 = 66 %) unterschieden werden. Lediglich bei den Reaktionszeiten ist zu sehen, dass die Probanden-Reaktionen stärker streuen und eine marginale Tendenz zu längeren Antwortzeiten im zweiten Szenario besteht. Daher könnte eine Änderung der Strategie vermutet werden, in welcher der Hauptaufgabe, also der Flugführung und der Bedienung der Navigationshilfe, im zweiten Szenario eine höhere Priorität eingeräumt wurde.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

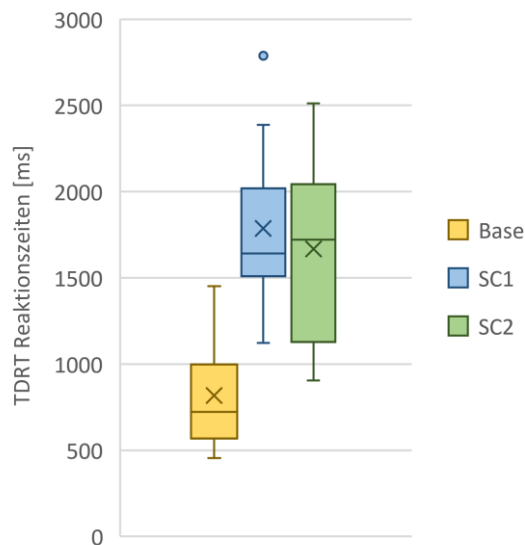


Abbildung 5.20 Boxplots der TDRT-Reaktionszeiten in Abhängigkeit der Szenario-Reihenfolge mit Darstellung der Baseline

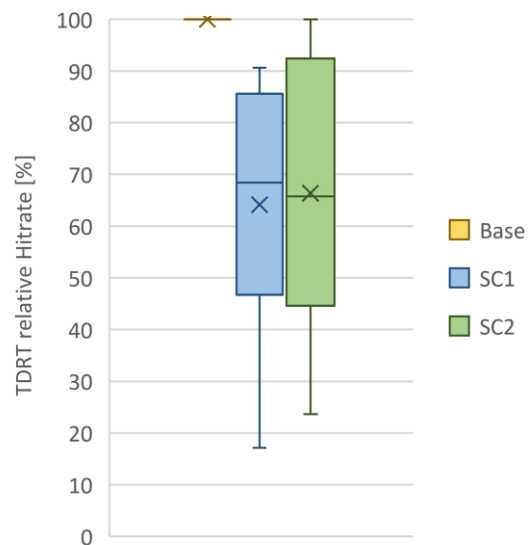


Abbildung 5.21 Boxplots der TDRT-Hit-Rates in Abhängigkeit der Szenario-Reihenfolge mit Darstellung der Baseline

Da dieses Vorgehen in der angewendeten Form nicht zu einer Differenzierung zwischen den beiden Szenarien geeignet war, scheint eine Anpassung der Methodik notwendig. Dies könnte zum einen in der Form gestaltet werden, dass die Reizschwelle minimiert wird, der gültige Antwortzeitrahmen verkleinert wird oder der dargebotene Reiz komplexer gestaltet wird und gegenüber einem Distraktor-Reiz im Rahmen der Signalentdeckungstheorie (s.a. Goldstein, 1999) erkannt werden muss. Gerade letzteres Vorgehen hat den Vorteil, dass die Methode an sich mehr mentale Kapazität einfordert und des Weiteren ein zu häufiges Reagieren auf Seiten der Probanden nicht ‚belohnt‘ wird.

Eye-Tracking

Ein weiteres Vorgehen zur objektiven Bewertung der Workload ist auch in diesem Kontext die Erhebung des Blickverhaltens. So wird eine Vergrößerung der Pupille mit einem Anstieg der Workload assoziiert (s.a. Kahneman, 1973; Manzey, 1998) und ein vermehrter Lidschluss von über 80 %, wie er beispielsweise anhand des PERCLOS erfasst wird, mit zunehmender Ermüdung (Wierwille et al., 1994). Es ist also zu erwarten, dass die Workload mit einer zunehmenden Sicherheit in der Bedienung der Navigationshilfe und auch im Hinblick von einer verbesserten Bekanntheit des Szenarien-Ablaufs sinkt, während der Grad an Ermüdung ansteigen sollte. Dies ist jedoch in diesem Fall weder bei Betrachtung der Pupillenerweiterung (Abbildung 5.22) noch im Hinblick auf das Lidschlussverhalten (Abbildung 5.23) der Fall. Der Pupillenerweiterungs-Median des ersten Szenarios (Mdn = 2257.46 px²) ist nahezu identisch mit dem Median des zweiten Szenarios (Mdn = 2384.18 px²). Zwar streuen die Daten im zweiten Szenario auch etwas mehr als im ersten Szenario, der Median des zweiten Szenarios ist aber wohl auch auf eine Tendenz zu höheren Werten zurückzuführen, da ein vergleichsweise kleiner Bereich das dritte Quartil bildet. Letztendlich ist anhand der Messung der Pupillenerweite-

5 Die Evaluierung des multidimensionalen Methoden-Inventars

rung jedoch keine Unterscheidung zwischen den beiden Szenarien möglich, die auf eine Verminderung der Workload im zweiten Szenario schließen lässt.

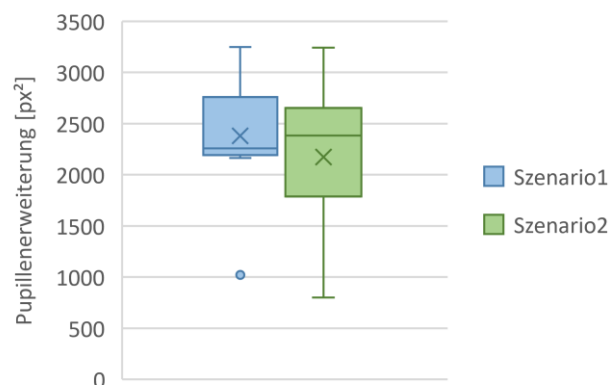


Abbildung 5.22 Boxplots der Pupillenerweiterung in Abhängigkeit der Szenario-Reihenfolge

Zur Bewertung der Ermüdung während der Bedienung der Navigationshilfe und der Szenario-Durchführung ist der PERCLOS verwendet worden, der laut Herstellerangaben der Firma Ergoneers den prozentualen Anteil des Lidschlusses über eine zu definierende Zeitspanne bemisst, welche hier die gesamte Szenario-Länge umfasste. Betrachtet man den PERCLOS über beide Szenarien, ist hier von keinem nennenswerten Unterschied auszugehen (Mdn Szenario 1 = 3.3 %; Mdn Szenario 2 = 2.8 %). Im Hinblick auf die Streuung der Daten zeigt sich im ersten Szenario sogar eher eine Tendenz zu höheren Ermüdungswerten als im zweiten Szenario.

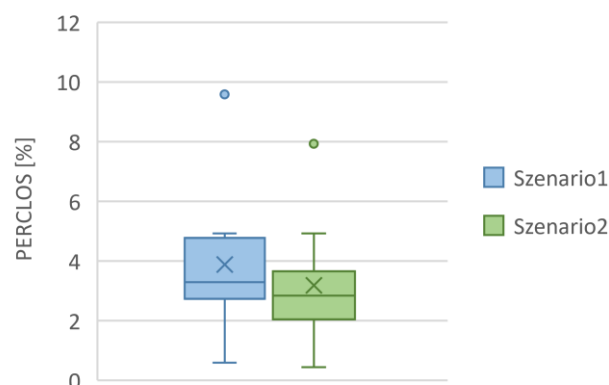


Abbildung 5.23 Boxplots des PERCLOS in Abhängigkeit der Szenario-Reihenfolge

Es stellt sich bei Analyse dieser Ergebnisse durchaus die Frage, welche Bedeutung diesen Werten zukommt. Zum einen könnte es sein, dass diese Messmethoden für den hier vorliegenden Anwendungsfall nicht sensitiv genug sind bzw. ein strenger kontrolliertes Laborsetting erfordern würden. Hinzu kommen aber auch Zweifel an den Daten per se. So könnte die Messumgebung selbst diese Daten – insbesondere die Pupillenerweiterung anhand der Lichtbedingungen – beeinflusst haben. Zudem stellt sich bei Betrachtung des PERCLOS und dessen Berechnung die Frage, ab wann von einer vermehrten Ermüdung auszugehen ist. Im Hinblick auf den hier vorliegenden PERCLOS-Daten ist

5 Die Evaluierung des multidimensionalen Methoden-Inventars

jedenfalls nicht von einer stärkeren Ermüdung der Probanden in einem der beiden Szenarien auszugehen.

Zusammenfassende Bewertung der Workload

Die Evaluierungsergebnisse und die verwendeten Methoden zur Erhebung der Workload unter Betrachtung der Bedienung einer re-designten Navigationshilfe unter Berücksichtigung der Szenarien-Reihenfolge vermitteln ein sehr ambivalentes Bild. Offensichtlich wird jedoch, dass ein subjektives Vorgehen wie beispielsweise der NASA TLX Fragebogen durchaus für Aussagen genutzt werden kann, die als Diskussionsgrundlage für weitere Design-Entscheidungen dienen können. Somit wird anhand des NASA TLX der plausible Rückschluss deutlich, dass das erste Szenario für die Probanden zu einer als stärker empfundenen Workload führt als das zweite Szenario, in welcher die Bedienung der Navigationshilfe noch stärker verinnerlicht werden konnte. Diese Unterscheidung wird bei den objektiven Daten-Erhebungen wie beispielsweise anhand des Eye-Trackings oder mittels TDRT nicht so deutlich. Die Verwendung der TDRT sollte weiter in deren Ausgestaltung – insbesondere in der Präsentation des Zielreizes – angepasst werden. Die Reaktionszeiten lassen hierbei zwar Aussagen zu dem Verhalten der Probanden im Cockpit zu, die Hit-Rate war hingegen nicht als Unterscheidungskriterium geeignet. Eine Auslegung im Sinne der Signalentdeckungstheorie wäre hier sinnvoll. Dies hätte insbesondere den Vorteil, dass diese Aufgabe nicht durch Probandenverhalten verfälscht werden kann. So könnte besonders sparsames Antworten oder übermäßiges Reagieren identifiziert werden und in der inhaltlichen Bewertung der Untersuchung berücksichtigt werden.

5.2.5.3 Bewertung der Gebrauchstauglichkeit im Rahmen der Evaluierung eines Re-Designs

Zur Analyse der Gebrauchstauglichkeit der re-designten Navigationshilfe sind als Nutzerleistung in Form von Bedienzeit sowie das Blickverhalten als objektive Kriterien gewählt worden. Der subjektive Aspekt der Nutzerzufriedenheit ist mit dem PSSUQ erhoben worden.

Nutzerleistung

Deutlich zu erkennen ist, dass die Probanden von einer häufigeren Bedienung der Navigationshilfe im Hinblick auf deren Eingabezeit profitieren. So wird ersichtlich, wie in Abbildung 5.24 dargestellt, dass die Probanden mit zunehmendem Training im zweiten Szenario, die Navigationspunkte schneller in das System eingeben können (Mdn = 16.77 s) als im ersten Szenario (Mdn = 26.67 s). Zudem nimmt die Streuung der Daten ab, d.h. es ist anzunehmen, dass die Piloten nahezu einheitlich von dieser Lernerfahrung profitieren.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

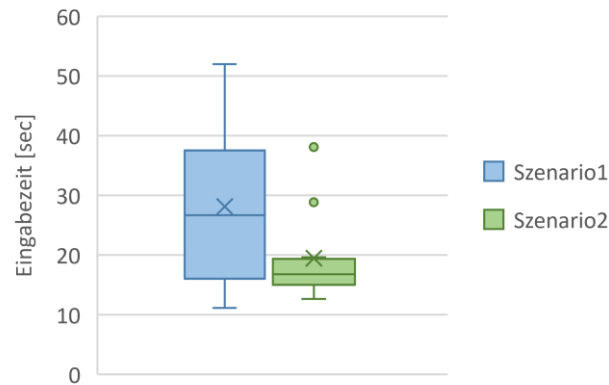


Abbildung 5.24 Boxplots der Eingabezeit in Abhängigkeit der Szenario-Reihenfolge

Post Study System Usability Scale (PSSUQ)

Die Ergebnisse der Nutzerzufriedenheit, die mittels PSSUQ erhoben worden sind, sind in Abbildung 5.25 zu finden. Der Wertebereich des positiv formulierten Fragebogens befindet sich in dem Bereich von 1 (stimme voll zu) bis 7 (stimme überhaupt nicht zu). Weiterhin hatten die Probanden die Möglichkeit, die Fragen mit freien Kommentaren zu ergänzen oder die Frage nicht zu beantworten. Eine Nichtbeantwortung des Fragebogens sollte allerdings begründet werden. Insgesamt zeigt sich, dass sich die Piloten relativ positiv hinsichtlich ihrer Nutzerzufriedenheit äußern, da sich die Mediane der drei PSSUQ-Faktoren sowie der Gesamtbewertung der Gebrauchstauglichkeit (Mittelwert über alle Fragen) im oberen Bereich der Werteskala befinden:

- Nützlichkeit des Systems (Mdn = 1.83)
- Informationsqualität (Mdn = 2.00)
- Qualität der Schnittstelle (Mdn = 2.33)
- Gebrauchstauglichkeit (Mdn = 1.91)

Lediglich die Informationsqualität – d.h. die Qualität dessen, was an Information dargestellt wird (Minimum = 1.17, Maximum = 3.67) und die Qualität der Schnittstelle, mit der beschrieben wird, wie und mit welchen Hilfsmitteln die darzustellende Information vermittelt wird (Minimum = 1.00, Maximum = 4.00), lassen noch Verbesserungsmöglichkeiten im Design vermuten, da hier die Daten etwas mehr streuen, was sich auch im Gesamtwert der Gebrauchstauglichkeit niederschlägt.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

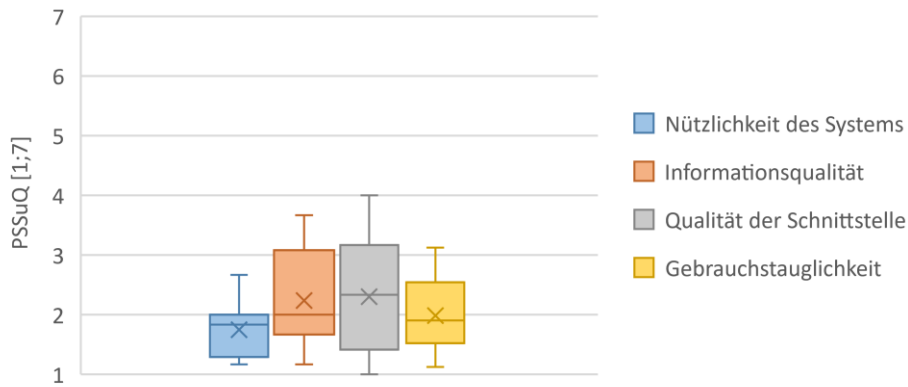


Abbildung 5.25 Boxplots der PSSUQ-Faktoren

Der PSSUQ ist in der Evaluierung der re-designten Navigationshilfe (RD) nur einmal verwendet worden, da nur eine technische Funktion in zwei verschiedenen Situationen bewertet wurde. Zur Ermittlung einer Vergleichsbasis wurden die Ergebnisse dieser Erhebung den Medianen aus der Machbarkeitsstudie (MS) wie in Kapitel 5.1 beschrieben und in Tabelle 5.3 und in Abbildung 5.26 gegenübergestellt.

Tabelle 5.3 Gegenüberstellung der PSSUQ-Mediane der Machbarkeitsstudie (MS) und der Re-Design-Studie (RD)

Design	Nützlichkeit des Systems	Informationsqualität	Qualität der Schnittstelle	Gebrauchstauglichkeit
MS alt	4.83	4.50	6.17	4.84
MS neu	1.83	2.42	2.33	2.16
RD	1.83	2.00	2.33	1.91

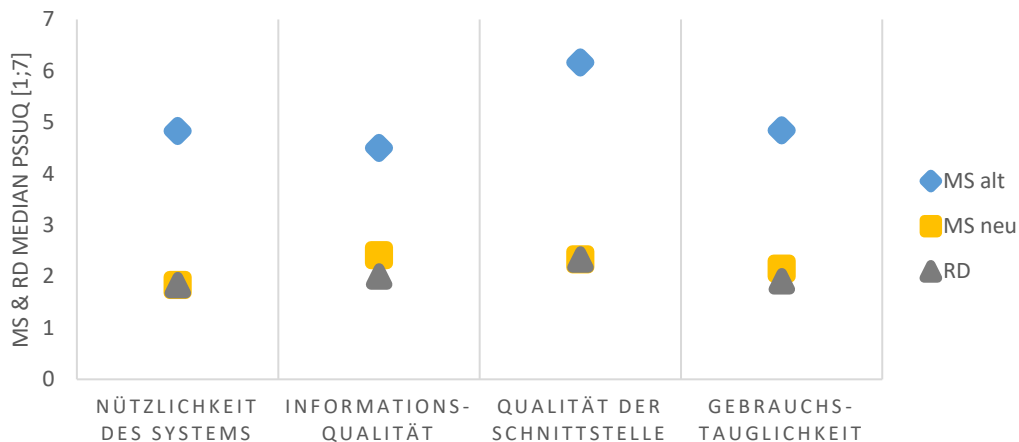


Abbildung 5.26 Gegenüberstellung der PSSUQ-Mediane der Machbarkeitsstudie (MS) und der Re-Design-Studie (RD)

Anmerkung: MS altes Design = blaue Rauten, MS 1. Iteration des Designs = gelbe Quadrate, im Vergleich zum RD als 2. Iteration des Designs = graue Dreiecke

Zu erkennen ist, dass die ursprüngliche Cockpit-Auslegung ohne Navigationshilfe zu deutlich schlechteren PSSUQ Urteilen in der Machbarkeitsuntersuchung geführt hat als die erste und zweite Iteration des neuen Designs. Im Re-Design ist lediglich eine marginale Verbesserung bei der Informa-

tionsqualität (Mdn = 2.00) im Vergleich zur ersten Iteration der Navigationshilfe (2.42) zu erkennen, die sich auch im errechneten Gesamtwert der Gebrauchstauglichkeit bzw. respektive der Nutzerzufriedenheit niederschlägt (Mdn RD = 1.91, Mdn MS neu = 2.16). Es wird somit erkennbar, dass eine weitere Iteration der Navigationshilfe nur noch einen geringen Mehrwert mit sich führen würde.

Eye-Tracking

Auch bei der Bewertung der Gebrauchstauglichkeit ist das Blickverhalten als objektives Vorgehen genutzt worden. Insbesondere ist hier die Verwendung von Sakkaden und Transitionen gemeint, da diese ein Indiz für die Verteilung von Informationen und somit der Suchstrategie der Probanden darstellen (Goldberg & Kotval, 1998). Transitionen im vorliegenden Vorgehen sind von Sakkaden insofern abzugrenzen, als dass sie orts- und somit inhaltsbezogen sind; da sie den Übergang von einem durch den Analysten definierten Bereich (AOI = Area of Interest) zum nächsten AOI, kennzeichnet (vgl. DIN EN ISO 15007-1:2003-10; Technical Specification ISO/ TS 15007-2:2001(E)). Sakkaden hingegen bezeichnen den Übergang von einer Fixation zur nächsten (s.a. Bojko, 2013; Holmqvist et al., 2011). Betrachtet man lediglich die prozentuale Transitionszeit (Abbildung 5.27) innerhalb der aufeinanderfolgenden Szenarien, ist von einem geringen Unterschied zum Nachteil des zweiten Szenarios auszugehen (Mdn Szenario 1 = 36 %, Mdn Szenario 2 = 41 %).

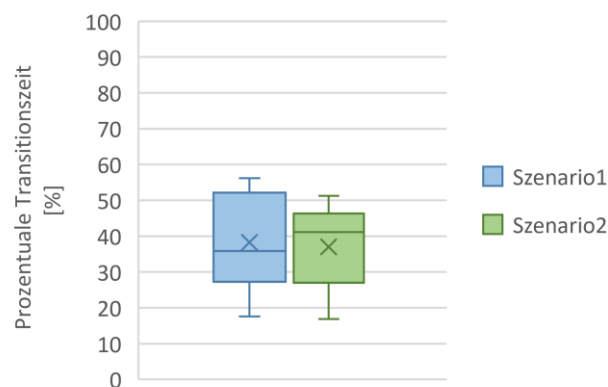


Abbildung 5.27 Boxplots der prozentualen Transitionszeit in Abhängigkeit der Szenario-Reihenfolge

Bei den Sakkaden hingegen ist ein leichter Vorteil für das zweite Szenarios zu erkennen, da hier möglicherweise eine Tendenz zu weniger Sakkaden besteht (Mdn Szenario 1 = 2063, Mdn Szenario 2 = 1676), die Probanden also zielgerichteter nach Informationen suchen (Abbildung 5.28). Ein Unterschied in der Dauer (Abbildung 5.29) der Sakkaden lässt sich augenscheinlich allerdings nicht feststellen (Mdn Szenario 1 = 28.41 ms, Mdn Szenario 2 = 27.54 ms).

5 Die Evaluierung des multidimensionalen Methoden-Inventars

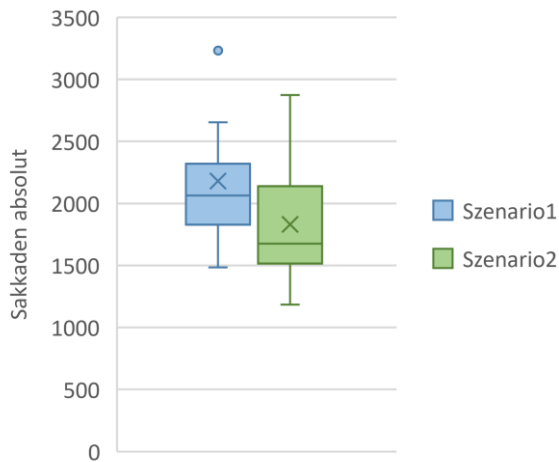


Abbildung 5.28 Boxplots der Anzahl an Sakkaden in Abhängigkeit der Szenario-Reihenfolge

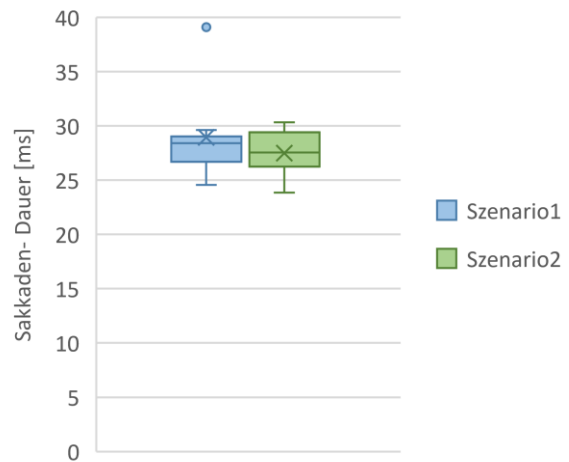


Abbildung 5.29 Boxplots der Sakkaden-Dauer in Abhängigkeit der Szenario-Reihenfolge

Etwas deutlicher wird der Zugewinn an Effizienz, wenn die Sakkaden spezifisch während der Bedienung der Navigationshilfe überprüft werden. In Abbildung 5.30 ist dargestellt, dass im zweiten Szenario weniger Sakkaden (Mdn = 372.50) auftreten als im ersten Szenario (Mdn = 523.00). Bei der Dauer der Sakkaden (Abbildung 5.31) zeigt sich allerdings erneut ein nicht weiter nennenswerter Unterschied von 29.28 ms im zweiten Szenario im Vergleich zum Median der Bedienung der Navigationshilfe im ersten Szenario mit 26.77 ms. Es ist somit zu vermuten, dass die Suchstrategie der Probanden im zweiten Szenario effizienter ist und sich dies in einer geringeren Anzahl an Sakkaden, die möglicherweise zielgerichteter sind, niederschlägt.

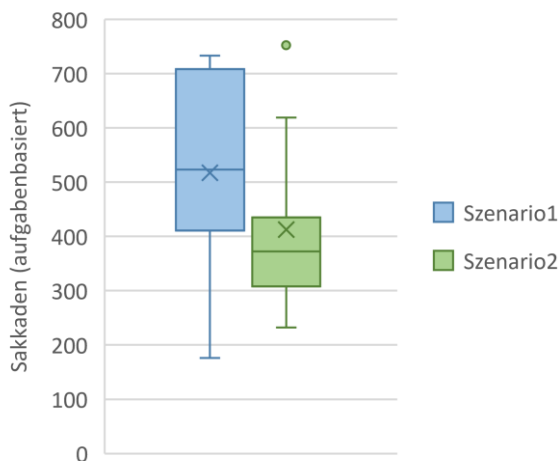


Abbildung 5.30 Boxplots der Anzahl an Sakkaden während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge

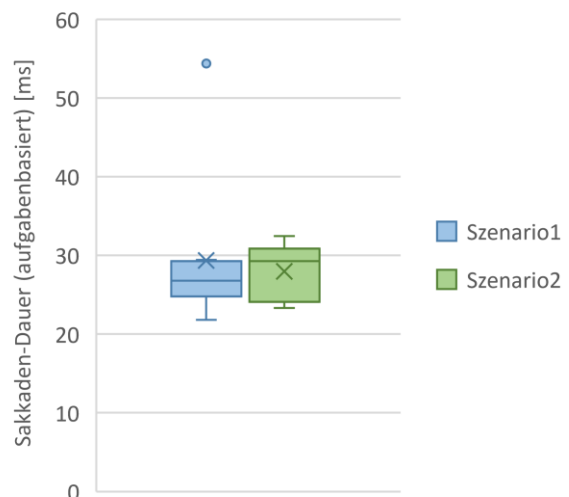


Abbildung 5.31 Boxplots der Sakkaden-Dauer während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge

Ergänzende qualitative Methoden

Zusätzlich zu den beschriebenen Methoden und deren Ergebnisse wurden noch Interviews am Cockpit und ein Cognitive Walkthrough im Rahmen einer Think-Aloud-Technik zur Erfassung der freien Gedanken während der Nutzung des Systems und zur Beschreibung der Gebrauchstauglichkeit der re-designten Navigationshilfe durchgeführt. Da beide Vorgehen insbesondere die inhaltliche Bewertung der Informationsdarstellung der Navigationshilfe erfassen und nicht weiter in der Validierungsmatrix verwendet werden, wird auf eine weitere Darstellung dieser Ergebnisse verzichtet.

Zusammenfassende Betrachtung der Gebrauchstauglichkeit

Bei Betrachtung der Gebrauchstauglichkeitsvariablen zur Untersuchung des Re-Designs stellt sich heraus, dass die Probanden von weiteren Lernerfahrungen mit der Navigationshilfe durch zusätzliches Training profitieren – die Eingabezeit also kürzer wird und das visuelle Blickverhalten effizienter. Befragt nach der Nutzerzufriedenheit geben die Probanden relativ positive Rückmeldungen bezüglich der zweiten Iteration der Navigationshilfe, was sich insbesondere im Vergleich zu den PSSUQ-Bewertungen aus der Machbarkeitsstudie zeigt. Da sich hier aber die Werte im Vergleich zur ersten Iteration nur wenig unterscheiden, würde eine weitere Designoptimierung nur keinen weiteren Mehrwert bieten.

Die Methoden-Kombination zur Bestimmung der Gebrauchstauglichkeit lässt einen geeigneten Einblick auf die Nutzerzufriedenheit und Effizienz der Bedienung zu. In diesem Zusammenhang scheint sich die Auswahl an subjektiven und objektiven Verfahren in ihrer Aussagekraft zu ergänzen. Für eine vollständige Bewertung der Gebrauchstauglichkeit sollte jedoch weiterhin ein geeignetes Verfahren in das Methoden-Inventar integriert werden, welches auch die Bedienfehler des Nutzers zuverlässig erfasst, um Aussagen bezüglich deren Eingabegenauigkeit zu treffen. Dies wurde hier zwar mittels eines Beobachtungsbogens versucht, konnte jedoch auf Grund des Umfangs des Szenarios nicht zuverlässig und objektiv bewertet werden.

5.2.5.4 Zusammenfassende Bewertung zur Untersuchung des Re-Designs einer Navigationshilfe

In einer ersten Überprüfung der Ergebnisse stellt sich heraus, dass die Schwierigkeit des Szenarios keinen Einfluss auf die Bedienung des Re-Designs der Navigationshilfe hat. Dies ist im Hinblick auf die Gebrauchstauglichkeit positiv zu bemerken, da die Navigationshilfe sowohl in Situationen mit einer geringeren Beanspruchung als auch in Situationen, in denen der Pilot unter stärkerer Arbeitsbelastung steht, gut genutzt werden kann. Im Nachgang sind die Daten zudem dahingehend zu interpretieren, inwiefern das Trainieren der Bedienung sich auf die Ausprägung der Qualitätskriterien auswirkt. Dahingehend sind die erhobenen Daten im Hinblick auf die Szenario-Reihenfolge – also ein geflogenes Szenario im Vergleich zu zwei geflogenen Szenarien zu interpretieren. Dies führt zu der Erkenntnis, dass eine häufigere Dateneingabe mit der Navigationshilfe zu einer Verbesserung der Bewertung hinsichtlich Gebrauchstauglichkeit, einer weniger stark empfundenen Workload sowie einem erhöhten Situationsbewusstsein auf Seiten der Probanden geführt hat.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Mehr Training, dass mit einer verbesserten Bedienung der Navigationshilfe einhergeht, führt dazu, dass die Probanden angeben, mehr Aufmerksamkeit auf die situativen Gegebenheiten richten zu können. Dies wird auch in den Blickdaten deutlich, da sich die Informationsaufnahme während der Bedienung der Navigationshilfe im zweiten Szenario als effizienter erweist. Die Probanden nehmen die Informationen gezielter, d.h. anhand weniger Fixationen, die dafür aber länger andauern, auf. Für ein verbessertes Situationsbewusstsein spricht auch, dass die räumliche Erinnerungsleistung im zweiten Szenario als höher einzustufen ist. Die Sortieraufgabe zur Erfassung der Handlungsstrategie scheint keinen Unterschied zwischen den beiden Szenarien zu kontrastieren. Allerdings betonen die Piloten in den Interviews, dass sie von einer häufigeren Bedienung profitiert haben, insbesondere da sie während der Szenarien eine Handlungsstrategie entwickelt hätten.

Die subjektiv wahrgenommene Workload wird von den Probanden über alle Dimensionen des NASA TLX im zweiten Szenario als geringer empfunden. Da die Daten jedoch eine hohe Streubreite aufweisen, ist davon auszugehen, dass die Piloten unterschiedlich stark von einer häufigeren Bedienung der Navigationshilfe profitieren. Dies ist möglicherweise mit dem unterschiedlichen Erfahrungshintergrund der Versuchspersonen zu begründen. Dass eine verbesserte Bedienung zu einer geringeren Workload führt, wird von den im Weiteren verwendeten Methoden nicht gestützt aber auch nicht widerlegt, da die in den Boxplots skizzierten Unterschiede zu gering sind.

Deutlich wird hingegen, dass die Piloten die Navigationshilfe nach häufigerer Bedienung schneller nutzen können. Auch ihre visuelle Suchstrategie wird effizienter, da die Anzahl an Sakkaden geringer wird. Der PSSUQ lässt insgesamt auf eine positive Bilanz der Nutzerzufriedenheit schließen. Bei Betrachtung der Subkategorien Nützlichkeit des Systems, Informationsqualität sowie der Qualität der Schnittstelle und der allgemein wahrgenommenen Nutzerzufriedenheit zeigt sich, dass alle Kategorien von den Nutzern als sehr positiv empfunden werden. Lediglich bei der Qualität der Schnittstelle deutet die Streubreite der Daten darauf hin, dass hier noch Verbesserungspotential vorliegt. Bei einem Vergleich mit der vorangegangenen Evaluierung des Erstdesigns der Navigationshilfe und der vorhergehenden Cockpit-Auslegung ohne Navigationshilfe stellt sich für das Re-Design heraus, dass sich nur die Informationsqualität und der Gesamtwert der Nutzerzufriedenheit geringfügig verbessert haben. Eine weitere Designoptimierung würde daher nur noch einen geringen Zugewinn bedeuten.

5.2.6 Bewertung der verwendeten Methoden-Kombination

Bei Evaluierung der Navigationshilfe stellt sich heraus, dass eine Methoden-Kombination aus verschiedenen subjektiven und objektiven Methoden zur Bewertung der Nutzerschnittstelle notwendig ist, um die Qualitätskriterien zu erfassen und die Vielschichtigkeit der Ergebnisse zu skizzieren. Der 3D SART hat sich in diesem Kontext als besonders einfaches Mittel herausgestellt, da seine schnelle Anwendbarkeit und die leicht verständliche Fragestellung sich gut in die Simulator-Studie einbinden lassen. Etwas schwieriger verhält es sich bei der Anwendung des angepassten SAGAT. Aufgrund der unterschiedlichen Fragestellungen wie der Stummen Karte, der verwendeten Sicherheitsurteile, der Identifikation der relevanten Wegpunkte sowie der Erfragung von relevanten Szenario-Ereignissen kann nicht sichergestellt werden, dass die jeweiligen Datenpunkte auch vergleichbar sind. Insbeson-

dere nach Betrachtung der Ergebnisse stellt sich zudem heraus, dass manche Fragen zu leicht waren, um Unterschiede zu kontrastieren. Die Sortieraufgabe als funktionsrelevantes Maß ist als methodisches Vorgehen leicht zu integrieren, aber in der Auswertung stichprobenabhängig und benötigt ein festes Regelwerk, nach der die Reihenfolge für die Handlungsabfolge festgelegt wird. Bezüglich der Erhebung des Blickverhaltens stellt sich die Anwendung auch in Bezug auf die Daten-Erhebung als aufwendig heraus, da von vornherein durch Kalibrierung und durch das Sicherstellen einer stabilen Beleuchtung Störungen vermieden werden müssen. Da die Experten-Stichproben naturgegeben klein sind, können des Weiteren auch keine Brillenträger ausgeschlossen werden, so dass auch hier ein stabiles System notwendig ist. Die Datenaufbereitung ist zudem sehr zeitaufwendig, da hier sichergestellt werden muss, dass die Daten ohne verbleibende Artefakte (beispielsweise durch Fehldetektionen der Pupille oder AOIs) verarbeitet werden. Der Aufwand ist unter Betrachtung des Situationsbewusstseins und der Gebrauchstauglichkeit durchaus gerechtfertigt, im vorliegenden Fall aber nicht zur Bemessung der Workload. Der NASA TLX zeichnet sich ebenso wie der 3D SART als sensitives Instrument aus, dass obwohl subjektiv schnell anzuwenden ist und sich zudem als gute Diskussionsgrundlage für die Erfassung des Situationsbewusstseins – respektive der Workload beispielsweise durch ein ergänzendes Interview eignet. Die Anwendung der TDRT scheint vielversprechend, da sie zu objektiv belastbaren Ergebnissen führt, muss aber insbesondere bei langen Szenarien, wie im vorliegenden Fall, angepasst werden – beispielsweise im Sinne der Signalentdeckungstheorie und unter einer verbesserten zufälligen Applikation der Reize und im Hinblick auf die Norm ISO 17488:2016. Zur Beschreibung der Gebrauchstauglichkeit in einer Simulator-Studie stellt sich hier heraus, dass Daten – wie beispielsweise die Bedienzeiten der Probanden – schnell abgegriffen und ausgewertet werden können, um dann auch zu belastbaren Aussagen zu führen. Die Anwendung des PSSUQ wäre hier allerdings ohne die entsprechende Vergleichsbasis aus der vorhergehenden Evaluierung des Erstdesigns der Navigationshilfe wenig aussagekräftig. Dennoch bieten die einzelnen Fragestellungen und die Möglichkeit zur Erfassung von freien Kommentaren eine solide Diskussionsgrundlage zur Erfassung der Gebrauchstauglichkeit. Neben den benannten Methoden wurden auch noch ein halbstrukturiertes Interview und ein Cognitive Walkthrough angewendet. Beide Vorgehen erweisen sich in der Auswertung als zeitaufwendig, die jedoch zu einem sehr vielschichtigen Ergebnisbild führen und auf die auch für die weitere Designentwicklung nicht verzichtet werden sollte. Die Verwendung dieser beiden Methoden zur Evaluierung der Navigationshilfe war jedoch zu umfangreich. Eine separate Erhebung der Aspekte, die nicht innerhalb Szenarien abgedeckt werden können, wäre daher zielführender gewesen. Erneut wird in der Quintessenz deutlich, dass eine vielschichtige Zusammensetzung von Methoden zu einer geeigneten Beschreibung der Qualitätskriterien führen und somit eine Bewertung der Cockpit-Modifikation ermöglicht.

5.3 Untersuchung eines räumlich-auditiven Warnsignals

Im Rahmen einer Untersuchung von Rabl et al. (2016) zur Evaluation eines räumlich-auditiv präsentierten Warnsignals im Cockpit, basierend auf einem System der Firma Terma, sind die Methoden des multidimensionalen Methoden-Inventars ebenfalls verwendet worden. Anhand einer kontrollierten Cockpit-Simulator-Studie ist überprüft worden, inwiefern ein auditiv-räumlich präsentiertes Warn-

5 Die Evaluierung des multidimensionalen Methoden-Inventars

signal mit ansteigender Dringlichkeit im Vergleich zur herkömmlichen Cockpit-Auslegung, die ein Warnsignal mit gleichbleibender Dringlichkeit und ohne auditiv-räumlichen Hinweis mit sich bringt, die Leistungen des Nutzers unterstützt. Um zu testen, wie performant die beiden Cockpit-Auslegungen und deren Designs unter zusätzlicher Belastung sind, haben die Autoren ergänzend eine taktile Zusatzaufgabe in das Versuchsdesign (siehe Tabelle 5.5) integriert.

5.3.1 Stichprobenbeschreibung

Es sind zehn männliche Nutzer mit einem Erfahrungshintergrund von 2775.5 Flugstunden im Mittel (Rabl et al., 2016) untersucht worden. Alle Probanden sind im Hinblick auf die Weiterverarbeitung und Verwendung ihrer Daten instruiert worden und nahmen freiwillig an der Untersuchung teil.

5.3.2 Verwendete Methoden

Zur Analyse des räumlich-auditiven Warnsignals sind die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit anhand verschiedener subjektiver wie auch objektiver Methoden aus dem Methoden-Inventar untersucht worden. Zur Ermittlung der Auswirkungen des räumlich-auditiven Warnsignals ist zudem der 3D SART zur subjektiven Befragung eingesetzt worden. Ergänzend hierzu wurde der SAGAT als weitere Methode mit subjektiven und objektiven Messelementen angewendet. Umgesetzt wurde der SAGAT anhand von aufgabenrelevanten Fragen zum Situationsbewusstsein, die randomisiert ausgewählt wurden und den Probanden nach Ablauf eines Trials gestellt wurden. In diesem Zusammenhang ist auch die Zeit, die die Versuchspersonen zur Beantwortung der Frage benötigten, miterfasst worden. Als weitere objektive Verfahren sind die Erfassung des Blickverhaltens sowie die Leistungen in der Gefahren-Detektierung zu sehen. Bezüglich der Gefahren-Detektierung ist erhoben worden, inwiefern die Probanden die Bedrohung, der sie ausgesetzt waren, richtig erkannten und adäquat handelten. Die TDRT ist als Zusatzaufgabe in das Versuchsdesign integriert worden und wird somit als unabhängige Variable eingesetzt – was dem Vergleich von Design-Alternativen in verschiedenen Situationen anhand von Belastung entspricht. Subjektiv erhoben wurde die Workload ergänzend mit dem NASA TLX und objektiv durch die Erfassung von Blickverhalten. Zur Beschreibung der Gebrauchstauglichkeit sind hauptsächlich objektive Messkriterien genutzt worden und die (Flug-) Leistungsdaten der Nutzer wurden aus der Simulation abgeleitet. Maße des Blickverhaltens wurden genutzt, um die Gebrauchstauglichkeit der Cockpit-Auslegungen zu erheben. Außerdem sind die Piloten jeweils zu ihrer Handlungsstrategie im Rahmen eines Kurzinterviews direkt am Cockpit-Simulator befragt worden, um den subjektiv empfundenen Einfluss des räumlich-auditiven Warnsignals zu überprüfen. Ein Überblick, welche Methoden im Rahmen der Untersuchung eines räumlich-auditiven Warnsignals verwendet wurden, ist in Tabelle 5.4 zu finden.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Tabelle 5.4 Methoden-Inventar zur Erhebung der Qualitätskriterien in einer Studie zur Evaluierung eines räumlich-auditiven Warnsignals

Methoden	Situationsbewusstsein	Workload	Gebrauchstauglichkeit
	3D SART SAGAT Funktionsrelevantes Maß: Gefahren-Detektierung Eye-Tracking	NASA TLX TDRT Eye-Tracking	Nutzerleistung Eye-Tracking Interviews

Anmerkung: Die qualitativen Methoden (Interviews) sind nicht Teil der ab Kapitel 6 beschriebenen Validierung

5.3.3 Szenario-Beschreibung

Es ist überprüft worden, inwiefern die räumliche Verortungsleistung der Probanden bei zwei verschiedenen Displays, das eine mit ansteigender Dringlichkeit und räumlich verortbaren auditiven Hinweisreiz, das andere mit gleichbleibender Dringlichkeit und ohne räumlich-auditiven Hinweisen durch eine Zusatzaufgabe beeinträchtigt werden. Hieraus ergibt sich ein 2 (Display) x2 (Zusatzaufgabe) Within-Versuchsdesign nach Rabl et al., 2016 wie in Tabelle 5.5 skizziert.

Tabelle 5.5 Versuchsplan nach Rabl et al., 2016 zur Evaluierung eines räumlich-auditiven Warnsignals bei unterschiedlicher Belastung

		Cockpit-Auslegung	
		Auditives Warnsignal	Räumlich-auditives Warnsignal
Zusatzaufgabe	Mit TDRT	Auditives Warnsignal mit TDRT	Räumlich-auditives Warnsignal mit TDRT
	Ohne TDRT	Auditives Warnsignal ohne TDRT	Räumlich-auditives Warnsignal ohne TDRT

Aufgabe des Piloten war es, während eines Fluges, in welchem Höhe, Richtung und Geschwindigkeit vorgegeben waren, auf eine im Cockpit-Simulator präsentierte Bedrohung, die dem Probanden anhand eines auditiven Warnsignals mitgeteilt wurde, mit einem vorgegebenen Flugmanöver zu reagieren. Dieses Flugmanöver sollte mit größtmöglicher Präzision durchgeführt werden, um der Bedrohung zu entgehen. Das Warnsignal in der bestehenden Cockpit-Auslegung ist auditiv und mit gleichbleibender Dringlichkeit sowie ohne Richtungshinweis und ist mit einem artifiziellen visuellen Reiz kombiniert worden. Auch bei der neuen Cockpit-Auslegung hat der Pilot neben dem auditiven Warnreiz, der in dieser Bedingung räumlich separiert und mit Dringlichkeitshinweis induziert wurde, einen artifiziellen, visuellen Reiz zur Orientierung erhalten. Wie oben beschrieben, ist des Weiteren die TDRT als Zusatzaufgabe und unabhängige Variable in das Versuchsdesign mit aufgenommen worden, welche die Cockpit-Auslegungen unter unterschiedlicher Belastung kontrastieren sollte. Ein Szenario repräsentierte somit eine Bedingung mit jeweils zwölf Trails bzw. Flugaufgaben, die unter unterschiedlicher Belastung erhoben wurden.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

5.3.4 Versuchsablauf

Nachdem der Proband begrüßt worden war und sich mit der Durchführung des Versuches einverstanden erklärt hatte, erhielt der Pilot eine kurze technische Schulung hinsichtlich der beiden verschiedenen Cockpit-Auslegungen sowie Schulungen des methodischen Vorgehens und des Versuchsablaufes. Anschließend trainierte der Proband das zu fliegende Manöver in beiden Cockpit-Auslegungen. Danach wurde die Leistung der Probanden in der Zweitaufgabe als Baseline-Messung erfasst und die Kalibrierung des Eye-Trackers vorgenommen, bevor die Durchführung des Experiments in allen randomisierten Bedingungen starten konnte. Während der Durchführung der Aufgabe wurden die Nutzerleistungen als Flugleistung, Reaktionen auf die TDRT und das Blickverhalten des Probanden erfasst. Nach jedem Trial erhielt der Proband zudem Fragen im Rahmen des SAGAT und der Gefahren-Detektierung. Im Anschluss an jede Bedingung wurde der Proband weiterhin zu seinen Handlungsstrategien befragt und beantwortete die Fragen des NASA TLX und des 3D SART. Das Experiment wurde beendet und der Proband verabschiedet, nachdem alle Experimentalbedingungen erfolgreich erfasst worden waren.

5.3.5 Bewertung der Verwendung eines räumlich-auditiven Warnsignals

Das räumlich-auditive Warnsignal ist mit verschiedenen objektiven und subjektiven Methoden untersucht worden, um dessen Eignung im Hinblick auf die Qualitätskriterien festzustellen. Nach A. Rabl, H. Neujahr und Zimmer (2016) stellt sich heraus, dass die Cockpit-Auslegung keinen Einfluss auf die Durchführung des Flugmanövers hat. Dieses Ergebnis ist insbesondere im Hinblick auf die Gebrauchstauglichkeit nennenswert, da dies dafür spricht, dass für die spezielle Flugaufgabe keine Verbesserung mit dem neuen Design zu erwarten ist. Betrachtet man hingegen das qualitative Blickverhalten der Probanden unter dem Aspekt Situationsbewusstsein, zeigt sich eine Tendenz, dass die Piloten von dem räumlich-auditiven Warnsignal profitieren, da sie mehr Zeit haben, um die Bedrohung außerhalb des Cockpits zu lokalisieren. Dies spiegelt sich auch in der Gefahren-Detektierung wider, die mit dem neuen Warnsignal insbesondere unter Belastung verbessert ist. Auch subjektiv äußern sich die Piloten im 3D SART dahingehend, dass sie ein verbessertes Verständnis der Situation haben. Dieses Ergebnis wird auch von den Aussagen der Probanden im Rahmen des NASA TLX gestützt, da hier alle Dimensionen bessere Werte für das neue auditiv-räumliche Warnsignal aufweisen – insbesondere die signifikanten Dimensionen ‚Leistung‘ und ‚Frustrationslevel‘. Die TDRT deutet ergänzend, aufgrund minimaler Effekte, die jedoch nicht signifikant sind, auf eine verbesserte mentale Kapazität bei räumlich-auditiven Warnungen hin. Ein Ergebnis, dass sich bei Betrachtung der Performance Operating Characteristics, wie sie Eichinger (2011) einführt, bestätigt.

5.3.6 Bewertung der verwendeten Methoden-Kombination

Bei Betrachtung der Methodenanwendungen ist festzustellen, dass sich die Erhebung des Blickverhaltens in diesem Rahmen als zielführend herausstellt. Dies ist insbesondere deshalb der Fall, weil die untersuchten Displays und Anzeigen groß genug waren, um zu interpretierbaren Ergebnissen zu führen, da es hier zu keinen Überlappungen der Bewertungsflächen anhand des Eye-Trackings kom-

men konnte. Die subjektiven Fragebögen wie NASA TLX und der 3D SART können die objektiven Daten stützen und tragen somit zu einer stimmigen Evaluierung des räumlich-auditiven Warnsignals bei. Bei der Verwendung des SAGAT stellt sich heraus, dass einige Items für die Probanden leichter verfügbar waren als andere. Da die Leistungen jedoch für deren weitere Interpretation in eine Skala transferiert werden müssen, ist die Konsistenz der Daten hier fraglich. Die TDRT ist erneut in der Auslegung eingesetzt worden wie im Re-Design der Navigationshilfe. Es ist anzumerken, dass eine Anpassung dieser Zweitaufgabe im Sinne der Signalentdeckungstheorie die freie mentale Kapazität besser erheben könnte, da das Antwortverhalten der Probanden zielführender erhoben werden kann. Weiterhin wurden zur Beurteilung der Nutzerleistung Simulator-Daten abgeleitet, mit denen sich Flugmanöverleistungen bewerten lassen. Ableitung und Verarbeitung dieser Daten erwiesen sich als umfangreich. Zusammenfassend lässt sich feststellen, dass eine Kombination von subjektiven und objektiven Methoden in der Evaluierung einer Bedienkomponente unter Berücksichtigung der Qualitätskriterien zu einer vielschichtigen und umfassenden Bewertung führt.

5.4 Untersuchung zweier Bedienformen zur Einstellung eines Sensors

In das bestehende System ist eine Bedienmöglichkeit integriert worden, die sowohl eine Veränderung der Genauigkeit eines Sensors sowie die Einstellung von dessen Sensitivität ermöglicht. Hierfür standen zwei Designlösungen zur Verfügung, die jeweils hinsichtlich ihrer Gebrauchstauglichkeit, der Nutzerbeanspruchung sowie bezüglich ihrer Auswirkungen auf das Situationsbewusstsein des Piloten getestet worden sind. Die zu testenden Designs haben sich zum einen in ihrer Menüführung und zum anderen hinsichtlich ihrer Einbindung in das bereits bestehende System unterschieden.

5.4.1 Stichprobenbeschreibung

Untersucht wurden fünf männliche Probanden, die den Versuch in allen Bedingungen durchführten. Ein weiterer Pilot wurde trainiert, konnte jedoch auf Grund einer technischen Fehlfunktion in der Simulationsumgebung den Versuch nicht durchführen, nahm jedoch am Interview teil. Die Versuchspersonen verfügten über eine Flugerfahrung die im Mittel bei 3416.67 Flugstunden mit einer Standardabweichung von 899.81 Flugstunden lag. Die Teilnahme der Versuchspersonen war freiwillig und die Weiterverarbeitung der Daten wurde von den Probanden bewilligt.

5.4.2 Verwendete Methoden

Zur Erfassung der Gebrauchstauglichkeit wurde das Blickverhalten und die Bedienzeit als objektive Messungen für beide Designs erfasst und ein Interview hinsichtlich der Nutzbarkeit der Designs durchgeführt. Zur Überprüfung der Beanspruchung der Probanden während der Bedienung wurde das Blickverhalten der Probanden erhoben und der Workload-Fragebogen NASA TLX verwendet. Um das Situationsbewusstsein der Probanden messen zu können, wurde außerdem der 3D SART eingesetzt und ebenfalls durch die Erhebung von Blickdaten ergänzt. Alle verwendeten Methoden werden als Übersicht in Tabelle 5.6 aufgeführt.

5 Die Evaluierung des multidimensionalen Methoden-Inventars

Tabelle 5.6 Methoden-Inventar zur Erhebung der Qualitätskriterien in einer Studie zur Evaluierung der Bedienung eines Sensors

Methoden	Situationsbewusstsein	Workload	Gebrauchstauglichkeit
	3D SART Eye-Tracking	NASA TLX Eye-Tracking	Nutzerleistung Eye-Tracking Interviews

Anmerkung: Die qualitativen Methoden (Interviews) sind nicht Teil der ab Kapitel 6 beschriebenen Validierung

5.4.3 Szenario-Beschreibung

Jedes Design wurde einzeln in einem hierfür entwickelten Szenario getestet. Aufgabe des Probanden war es, während festgelegter Zeitpunkte im Szenario die Sensitivität des Sensors zu verstellen und gemäß den Ereignissen im Szenario auch die Genauigkeit des Sensors anzupassen. Der Pilot sollte hierzu eine vorgegebene, kreisförmige Route abfliegen, auf welcher jeweils in bestimmten ortsbezogenen Abschnitten Bedrohungen oder Anweisungen zur Anpassung der Sensitivität stattfanden. In 75 % aller Fälle erhielt der Proband zwischen der Anpassung der Sensitivität und der Anpassung der Genauigkeit des Sensors eine Zusatzaufgabe. Die Bearbeitung dieser zusätzlichen Aufgabe wurde zeitlich so abgestimmt, dass in 25 % der Fälle ausreichend Zeit bestand, diese zu erfüllen. In weiteren 25 % der Fälle konnte die Zusatzaufgabe gerade noch beendet werden und in weiteren 25 % wurde diese Zusatzaufgabe durch eine auftretende Bedrohung unterbrochen. In den restlichen 25 % der Fälle wurde keine Zusatzaufgabe gegeben.

5.4.4 Versuchsablauf

Die Nutzer erhielten vorab eine technische Beschreibung hinsichtlich beider Designlösungen. Diese wurden dem Probanden vor Beginn des Experimentes nochmals präsentiert und Rückfragen wurden beantwortet. Außerdem erhielt der Proband eine Beschreibung des genauen Versuchsablaufes sowie eine Übersicht der verwendeten Methoden mit zugehöriger Instruktion. Anschließend konnten die Probanden die Bedienung der beiden Designlösungen im Cockpit-Simulator erlernen und diese in den hierfür entwickelten Szenarien üben. Es wurde sichergestellt, dass alle Probanden ausreichend trainiert waren. Nach dem Training wurde die Kalibrierung des Eye-Trackings vorgenommen und der Proband konnte im Anschluss das Szenario mit Hinterlegung einer der beiden Bedienformen beginnen. Jedes Szenario bestand hierbei aus insgesamt zwölf vorzunehmenden Einstellungen. Nach Beendigung des Szenarios wurden ergänzend die Fragebögen 3D SART und der NASA TLX angewendet, bevor die nächste Trainingsphase startete. Nach Abschluss des zweiten Szenarios in der verbleibenden Bedienform wurde die Evaluierung nach der Durchführung eines halbstrukturierten Interviews abgeschlossen und der Proband wurde verabschiedet.

5.4.5 Bewertung der Bedienformen zur Einstellung eines Sensors

Unter Berücksichtigung der Ergebnisse aller verwendeten Methoden und der subjektiven Bewertung der Probanden im halbstrukturierten Interview stellt sich heraus, dass beide Bedienformen zwar akzeptabel sind, jedoch eine Bedienform bei den Nutzern etwas beliebter ist. Auch wenn die genutzt-

5 Die Evaluierung des multidimensionalen Methoden-Inventars

ten Methoden keinen zu betrachtenden Unterschied zwischen den beiden Bedienformen kontrastieren, zeigt sich doch ein marginaler Vorteil für das Bedienkonzept, welches auch die Nutzer favorisierten. Lediglich die Betrachtung der Bedienzeit zur Einstellung der Sensor-Sensitivität führt zu einer andersartigen Antwort, da in der nicht favorisierten Bedienform weniger sukzessive Bedienvorgänge notwendig sind.

5.4.6 Bewertung der verwendeten Methoden-Kombination

Im Hinblick auf die Bewertung der verwendeten Methoden wird bei dieser Evaluierung erneut deutlich, dass mehrere Methoden zur Bestimmung der Qualitätskriterien genutzt werden müssen. Diese sollten sich sinnvoll ergänzen, da im vorliegenden Kontext die verwendeten Methoden im Einzelnen gesehen nur eine geringe Aussagekraft haben. In der Gesamtschau und unter Berücksichtigung der Interviewergebnisse wird jedoch ein stimmiges Bild hinsichtlich der Favorisierung eines Bedienkonzeptes deutlich, dass auch für die Weiterentwicklung des Designs genutzt werden kann. Lediglich die Verwendung des Eye-Tracking Systems stellt sich bezüglich des Umfangs dieser Evaluierung als zu aufwändig bezüglich der Datenanalyse heraus, ohne einen nennenswerten Mehrwert hinsichtlich der Beurteilung der Bedienkonzepte zu liefern.

6 DER MULTI-TRAIT/MULTI-METHOD-ANSATZ ALS STRATEGIE ZUR ERHEBUNG DER KONSTRUKT-VALIDITÄT

Sind die verschiedenen Cockpit-Funktionen mit dem multidimensionalen Methoden-Inventar evaluiert, kann zum einen festgestellt werden, ob sich die verwendeten Methoden prinzipiell für die Erhebung der Qualitätskriterien eignen und zum anderen kann bewertet werden, ob diese Methoden speziell im Cockpit-Simulator anwendbar sind. Zur Beantwortung beider Fragen ist eine Validierungsstrategie unter Betrachtung der Konstrukt-Validität (vgl. Kapitel 6.1) wie beispielsweise mit Hilfe des Multi-Trait-Multi-Method Ansatzes (vgl. Kapitel 6.2) erforderlich.

6.1 Die Konstrukt-Validität als Grundlage einer Validierungsstrategie

Die Validität, als eines der drei Hauptgütekriterien, gibt die Genauigkeit eines Tests oder einer Methode an, mit der sie die zu beschreibende Verhaltensweise oder ein Persönlichkeitsmerkmal tatsächlich misst (Bortz & Döring, 2006). Nach Lienert (1998) werden drei Überbegriffe der Validität unterschieden, die wiederum nach Bühner (2011) mehrere Teilaspekte der Validität umschreiben: Inhaltliche Validität (1), Konstrukt-Validität (2) und die Kriteriumsvalidität (3). Die Inhaltsvalidität (1) als Begriff ist hierbei nur schwer fassbar, da sie empirisch nicht nachprüfbar ist, weil der Anspruch erhoben wird, dass die verwendeten Testitems das Konstrukt repräsentativ und ohne Überschneidungen zu anderen Konstrukten zuverlässig abbilden (Bühner, 2011). Unter inhaltlicher Validität wird demnach auch die Bewertung von Experten bezüglich der Eignung eines Tests als ‚Konsens von Kundigen‘ verstanden (Lienert & Raatz, 1998). Mit der Inhaltsvalidität einher gehen auch die Begriffe logische Validität und Augenscheinvalidität. Die logische Validität wird ähnlich dem Begriff der Inhaltsvalidität verwendet (Lienert & Raatz, 1998), während die Augenscheinvalidität einen offensichtlichen Zusammenhang von Test und Eigenschaft beschreibt (Bühner, 2011). Im Gegensatz zur Inhaltsvalidität ist die Kriteriumsvalidität (3) empirisch erhebbar, da sie den Zusammenhang eines latenten Testkennwertes mit einem Kriterium anhand eines manifesten Kennwertes beschreibt. Unterteilt wird die Kriteriumsvalidität nach Bühner (2011) in die prognostische Validität, die Übereinstimmungsvalidität, die retrospektive Validität und die inkrementelle Validität. Während die prognostische Validität anhand dessen bestimmt wird, ob ein Test zukünftiges Verhalten korrekt vorhersagen kann (Bortz & Döring, 2006), wird bei der retrospektiven Validität der Zusammenhang mittels eines zurückliegenden Kriteriums ermittelt (Bühner, 2011). Folglich werden bei der Übereinstimmungsvalidität Test-Wert und Kriteriums-Wert nahezu zeitgleich erhoben (Bortz & Döring, 2006). Die inkrementelle Validität als letzter Teilaspekt der Kriteriumsvalidität beschreibt hingegen den Zugewinn der Vorhersage eines Kriteriums über einen weiteren Test hinaus (Bühner, 2011).

Entscheidend für das weitere Vorgehen ist hingegen die Konstrukt-Validität (2). Diese lässt sich mit Hilfe der konvergenten Validität, der diskriminanten Validität und der faktorielle Validität beschreiben und bildet den mittleren Strang der nach Bühner (2011) visualisierten Validitätsformen und deren Teilaspekte (vgl. Abbildung 6.1 Validitätsformen und deren Teilaspekte nach Bühner (2011) Abbildung 6.1).

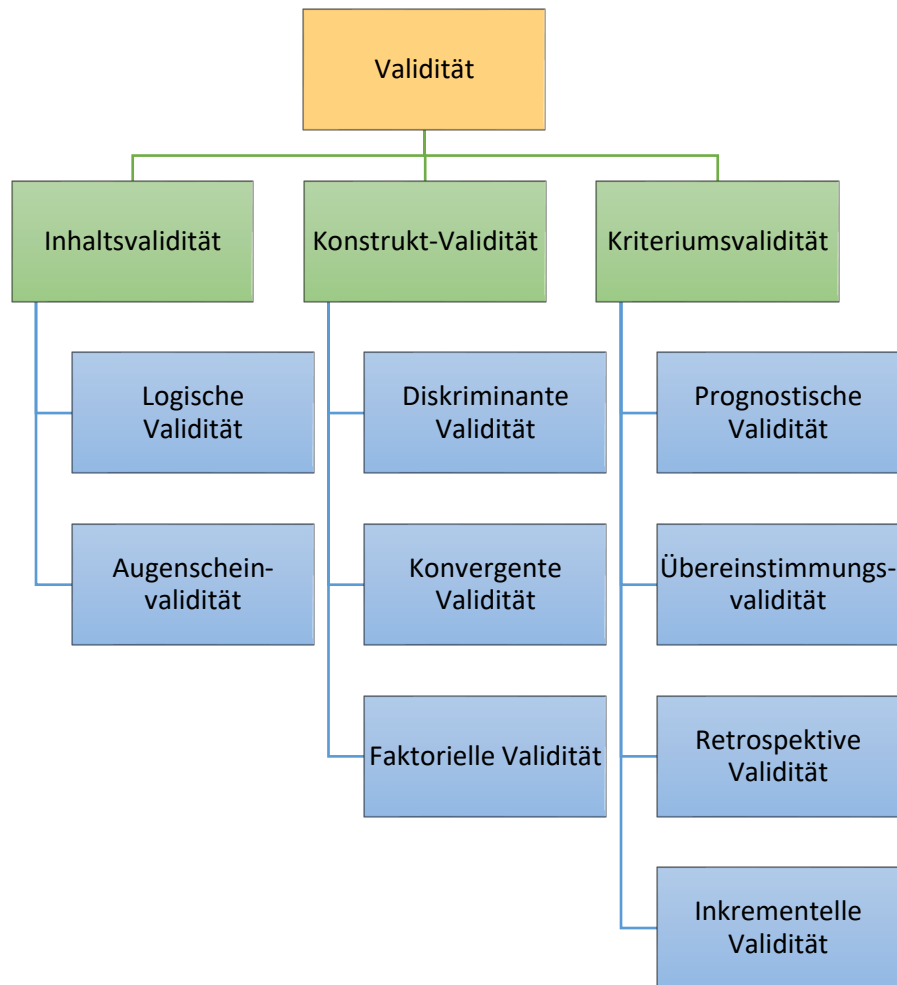


Abbildung 6.1 Validitätsformen und deren Teilaspekte nach Bühner (2011)

Eine Konstrukt-Validierung ist dann notwendig, wenn die verfügbaren operationalisierten Kriterien im Einzelnen nicht adäquat genug sind, um das zu messende Konzept zu beschreiben und hierfür mehrere indirekte Messungen vorzunehmen sind. Unter einem Konstrukt versteht man demnach ein postuliertes menschliches Attribut, von dem man ausgeht, dass es mit Hilfe eines Tests erfasst werden kann (Cronbach & Meehl, 1955). Zur experimentellen Untersuchung der Konstrukt-Validität schlagen Cronbach und Meehl (1955) verschiedene Vorgehen wie die Untersuchung nach Gruppenunterschieden, Untersuchungen zur internen Struktur eines Konstrukts, Prozessstudien, Untersuchungen über sich verändernde Situationen und die Analyse von Korrelationsmatrizen vor. Die Analyse von Korrelationsmatrizen zur Beschreibung von Konstrukt-Validität unter Betrachtung von Methodeneinflüssen haben hierbei Campbell und Fiske (1959) in Form des Multi-Trait/ Multi-Method Ansatz ins Leben gerufen und haben weiterhin den Begriff der konvergenten Validität um den der diskriminanten Validität erweitert (Schmermelleh-Engel & Schweizer, 2012). Unter konvergenter Validität versteht man hierbei, dass die erzielten Messergebnisse zu einem theoretischen Konstrukt von mehreren unabhängigen Methoden bestätigt werden. Die diskriminante Validität hingegen beschreibt die Abgrenzung der Messergebnisse zu anderen theoretischen Konstrukten (Campbell & Fiske, 1959). Auch die faktorielle Validität ist – rein logisch gesehen – ein Teilaspekt der Konstrukt-Validität und betrachtet die faktorielle Struktur eines Tests (Lienert & Raatz, 1998). Die Zugehörigkeit

der Testergebnisse werden hier ebenso bezüglich ihrer Konstrukt-Nähe kontrastiert und von konstruktfernen Bereichen getrennt (Bühner, 2011).

Da es sich bei den Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit um theoretische Konstrukte handelt, ist als Validierungsstrategie eine Konstrukt-Validierung der unter Kapitel 3.4 beschriebenen Methoden vorzunehmen. Als Ansatz wird hier der Multi-Trait/ Multi-Method-Ansatz gewählt, da er wie Hartig, Frey und Jude (2012) zur Konstrukt-Validierung fordern, experimentelle Effekte und korrelative Zusammenhänge mit anderen theoretischen Variablen betrachtet (Hartig, Frey & Jude, 2012). Da eine Analyse vieler Korrelationen zudem schnell unübersichtlich werden kann, bietet der Multi-Trait/ Multi-Method Ansatz eine systematische Strategie für deren Interpretation (Bühner, 2011).

6.2 Der Multi-Trait/Multi-Method-Ansatz als Validierungsstrategie

Die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit können nicht unabhängig voneinander betrachtet werden, da diese Kriterien gemeinsam anhand eines Evaluationsgegenstandes im Cockpit-Simulator erhoben werden und sich gegenseitig beeinflussen. Da die Qualitätskriterien jedoch jeweils als eigenständige Konstrukte zu verstehen sind, aber anhand eines Mixed-Method-Ansatzes, mit welchem ein Untersuchungsgegenstand anhand verschiedener Methoden und Daten sowie unterschiedlicher Theorien untersucht wird (Bortz & Döring, 2006), erhoben wurde, ist eine systematische Validierungsstrategie von Nöten.

Daher werden die Ergebnisse der vorgestellten Methoden des Methoden-Inventars dazu genutzt, sich gegenseitig zu ergänzen und damit zu präzisieren. Grundlegend hierfür ist der Gedanke, dass die Methoden, die das gleiche Qualitätskriterium messen, sich ähnlich verhalten sollten – also konvergent sind. Im Rückschluss sollten daher Methoden, die ein anderes Konstrukt oder Qualitätskriterium messen, sich voneinander unterscheiden (Campbell & Fiske, 1959). Ein strukturiertes Vorgehen dieser Validierungsstrategie wird durch die Anwendung des Multi-Trait/ Multi-Method-Ansatzes nach Campbell & Fiske (1959) möglich, welche sich bei der Verwendung mehrerer Methoden mit quantitativem Datenmaterial eignet. Das Vorgehen umfasst vier Aspekte, die zum Nachweis der Gültigkeit des Methoden-Inventars erfüllt werden müssen (Bühner, 2011; Campbell & Fiske, 1959):

1. Die Ergebnisse mehrerer voneinander unabhängige Methoden eines Merkmals müssen sich zueinander konvergent verhalten (Konvergenzkriterium).
2. Die verwendeten Methoden müssen sich ausreichend voneinander unterscheiden, sofern sie verschiedene Merkmale messen (Diskriminanz-Kriterium).
3. Jeder Test und jede Aufgabe der Messung ist eine eigenständige Trait-Methoden-Einheit.
4. Mehr als ein Merkmal und mehr als eine Methode sind für die Validierung notwendig.

In Abbildung 6.2 wird eine Matrix nach Campbell und Fiske dargestellt und exemplarisch auf den vorhandenen Kontext angewendet. Die Qualitätskriterien Situationsbewusstsein (SB), Workload (WL) und Gebrauchstauglichkeit (GT) oder in diesem Fall Traits, werden anhand von drei methodischen

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

Vorgehen – Selbstbeobachtung, Fremdbeobachtung und Leistungsmessung – erschlossen. Zur Validierung werden vier Aspekte der Matrix betrachtet:

- Die MonoTrait-MonoMethod-Diagonale (MTMM): Bezeichnenderweise trägt diese Diagonale auch den Namen Reliabilitätsdiagonale, da hier jeweils ein Trait mit ein und der selben Methode gemessen wird (Campbell & Fiske, 1959).
- Die HeteroTrait-MonoMethod-Dreiecke (HTMM): Verschiedene Merkmale werden mit einer Methode gemessen, d.h. der Einfluss einer Methode auf die unterschiedlichen Merkmale wird erfasst (Schmermelleh-Engel & Schweizer, 2012).
- Die MonoTrait-HeteroMethod-Diagonale (MTHM): Diese Diagonale wird auch als Validitätsdiagonale (Campbell & Fiske, 1959) bezeichnet, da hier ein Trait oder Merkmal mit mehreren Methoden erfasst wird und die Unterschiede der Korrelationen somit nur auf den Einsatz der verschiedenen Methoden rückführbar sind. Zusammen mit der MTMM-Diagonalen erfassen hier die Korrelationskoeffizienten die konvergierende Validität (Schmermelleh-Engel & Schweizer, 2012).
- HeteroTrait-HeteroMethod-Dreiecke (HTHM): Innerhalb der HTHM-Dreiecke – so fordert es das Diskriminanz-Kriterium – ist mit den geringsten Korrelationen zu rechnen, da die verschiedenen Traits mit für sie verschiedenen Methoden gemessen werden (Bühner, 2011).

Bei der Multi-Trait/ Multi-Method-Matrix ist also zu erwarten, dass die Korrelationen der MTMM-Diagonalen höher als die der MTHM-Diagonalen sind, die wiederum höher als die Korrelationen der HTMM-Dreiecke und der HTHM-Dreiecke sind (Schmermelleh-Engel & Schweizer, 2012).

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

		Selbstbeobachtung			Fremdbeobachtung			Leistungsmessung		
		SB	WL	GT	SB	WL	GT	SB	WL	GT
Selbstbeobachtung	SB	MTMM								
	WL	HTMM	MTMM							
	GT	HTMM	HTMM	MTMM						
Fremdbeobachtung	SB	MTHM	HTHM	HTHM	MTMM					
	WL	HTHM	MTHM	HTHM	HTMM	MTMM				
	GT	HTHM	HTHM	MTHM	HTMM	HTMM	MTMM			
Leistungsmessung	SB	MTHM	HTHM	HTHM	MTHM	HTHM	HTHM	MTMM		
	WL	HTHM	MTHM	HTHM	HTHM	MTHM	HTHM	HTMM	MTMM	
	GT	HTHM	HTHM	MTHM	HTHM	HTHM	MTHM	HTMM	HTMM	MTMM

Abbildung 6.2 Exemplarische Multi-Trait/ Multi-Method-Matrix angepasst nach Campbell und Fiske (1959)

Anmerkung: Abbildung 6.2 zeigt eine exemplarische Multi-Trait/ Multi-Method-Matrix zur Erhebung der Qualitätskriterien nach dem Ansatz von Campbell und Fiske (1959). Zu sehen sind die blaue Reliabilitätsdiagonale (MTMM), die HTMM-Dreiecke (grün), die orangen Validitätsdiagonalen (MTHM) und die grauen HTHM-Dreiecke, welche entstehen, wenn mehrere Traits durch verschiedene Methoden erfasst werden.

Der Ansatz für sich zielt darauf ab, dass die Varianz der Ergebnisse nicht nur auf die Unterschiede zwischen den Probanden zurückzuführen ist, sondern auch auf die Erhebungsmethode selbst (Bühner, 2011). Diskrepanzen in den verwendeten Messmethoden können somit beleuchtet und Hinweise zur Feinjustierung und Anpassung der Methoden aus den Ergebnissen extrahiert werden. Eine Überprüfung der Konstrukt-Validität anhand dieser strukturierten, jedoch aufwendigen Validierungsstrategie basiert auf einer visuellen Inspektion. Campbell und Fiske geben allerdings nicht vor, wie hoch die Korrelationen der konvergenten Validität sein müssten bzw. wie niedrig die Korrelationen der diskriminanten Validität sein sollten (Schmermelleh-Engel & Schweizer, 2012).

Bereits 2018 zeigen Billeriß C., Kellerer J. und Zimmer A., dass der MTMM-Ansatz für eine Validierung des Methoden-Inventars und im Rückschluss auch der Qualitätskriterien angepasst werden muss, da der dritte Validierungsprozess nach Campbell und Fiske, in welchem gefordert wird, dass jede Aufgabe der Messung eine eigenständige Trait-Methoden-Einheit ist, nicht erfüllt werden kann. Das ist laut Billeriß et. al. (2018) vor allem darin begründet, dass die Qualitätskriterien durch verschiedene Mess-Items und stark unterschiedlichen Methoden bestimmt werden, die nicht zusammengefasst werden können. Zwar wird die Matrix wie auch im obengenannten Beispiel durch die Qualitätskriterien bestimmt, die Methoden, die die Qualitätskriterien messen, sind jedoch durch unterschiedliche Mess-Items definiert. D.h. Fragebögen, psycho-physiologische Messungen und Leistungsmessungen werden so verschieden zur Messung der Qualitätskriterien eingesetzt, dass diese nicht als Strukturierung der Matrix dienen können.

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

Zur Bestimmung der globalen Evaluationsmatrix wird daher jedes Methodenitem mit jedem anderen Methodenitem korreliert und der Spearman Korrelationskoeffizient für jede Korrelation in die Matrix eingetragen. Die Matrix besteht somit aus Item-Korrelationen mit einer Diagonalen aus Eigenkorrelationen. Da die Matrix somit in der Mitte gespiegelt wird, ist eine Interpretation der Daten in einer horizontalen Leseweise vollständig und umfassend.

Der Aufbau der Evaluationsmatrix wird schematisch in Abbildung 6.3 dargestellt (Billeriß et al., 2018). Die Items und deren Korrelationen werden zu den ihnen zugehörigen Methoden gruppiert. Alle Methoden wiederum werden so zusammengefasst, dass diese jeweils das Qualitätskriterium bestimmen, das sie auch messen. Das heißt, dass die Qualitätskriterien durch die Methoden erfasst werden und die Methoden durch ihre Mess-Items. Folglich wird die konvergente Validität durch die Items und Methoden eines Qualitätskriteriums definiert und die diskriminante Validität durch die Korrelationen mit den anderen Qualitätskriterien. Zur Analyse der konvergenten Validität werden zwei Submatrizen betrachtet (vgl. Kapitel 6.2.1), die MonoMethod-MonoTrait-Matrix (MMMT) und die HeteroMethod-MonoTrait-Matrix (HMMT), die sich im jeweiligen Quadranten der Qualitätskriterien befinden (siehe hierzu beispielsweise den gelben Kasten zur Bestimmung der konvergenten Validität von Situationsbewusstsein in Abbildung 6.3). Als Traits werden in diesem Zusammenhang die jeweils zu überprüfenden Qualitätskriterien verstanden. Auch zur Untersuchung der diskriminanten Validität stehen zwei Submatrizen zur Verfügung, da drei Qualitätskriterien analysiert werden. Hierzu werden die Items eines Qualitätskriteriums mit den Items eines der beiden anderen Qualitätskriterien ermittelt, d.h. beispielsweise Situationsbewusstsein mit Workload und Situationsbewusstsein mit Gebrauchstauglichkeit. Diese beiden Submatrizen werden unter dem Begriff der HeteroMethod-HeteroTrait-Matrix (HMHT) zusammengefasst (siehe Kapitel 6.2.2).

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

			Situationsbewusstsein					
			SB			SB		
			Methode1		Methode2			
			SB Item 1	SB Item 2	SB Item 3	SB Item 4	SB Item 5	SB Item 6
Situationsbewusstsein	SB	Methode1	SB Item 1					
		Methode2	SB Item 2					
		SB Item 3						
	SB	Methode1	SB Item 4					
		Methode2	SB Item 5					
		SB Item 6						
			Workload					
			WL			WL		
			Methode1		Methode2			
			WL Item 1	WL Item 2	WL Item 3	WL Item 4	WL Item 5	WL Item 6
Workload	WL	Methode1	WL Item 1					
		Methode2	WL Item 2					
		WL Item 3						
	WL	Methode1	WL Item 4					
		Methode2	WL Item 5					
		WL Item 6						
			Gebrauchstauglichkeit					
			GT			GT		
			Methode1		Methode2			
			GT Item 1	GT Item 2	GT Item 3	GT Item 4	GT Item 5	GT Item 6
Gebrauchstauglichkeit	GT	Methode1	GT Item 1					
		Methode2	GT Item 2					
		GT Item 3						
	GT	Methode1	GT Item 4					
		Methode2	GT Item 5					
		GT Item 6						

Abbildung 6.3 Schematische Darstellung einer Evaluationsmatrix

Anmerkung: Schematische Darstellung einer Evaluationsmatrix nach Durchführung eines Experimentes (vgl. Billeri^ß et. al., 2018). Aufgrund der Qualitätskriterien ergeben sich drei Blöcke (Situationsbewusstsein gelb; Workload blau; Gebrauchstauglichkeit orange), die die Matrix gliedern. Jeder Block wird durch die zugehörigen Methoden und deren Items bestimmt. Die grüne Diagonale stellt die Eigenkorrelationen der Items mit sich dar. Die konvergente Validität kann anhand der Item-Korrelationen eines Qualitätskriteriums entlang der Diagonalen analysiert werden (Situationsbewusstsein gelbe Umrandung; Workload blaue Umrandung, Gebrauchstauglichkeit orange-braune Umrandung)

6.2.1 Sub-Matrizen zur Bestimmung konvergenter Validität

Zur Bestimmung der konvergenten Validität stehen zwei Sub-Matrizen zur Verfügung, die MonoMethod-MonoTrait-Matrix (MMMT-Matrix) und die HeteroMethod-MonoTrait-Matrix (HMMT-Matrix). Beide Sub-Matrizen sind eine fokussierte Sichtweise auf die globale Evaluationsmatrix, die sich nur innerhalb eines Qualitätskriteriums abspielt – also jeweils für Situationsbewusstsein, Workload und Gebrauchstauglichkeit getrennt.

Bei der MTMM-Matrix (vgl. Abbildung 6.4) wird nur eine Methode mitsamt der ihr immanenten Mess-Items betrachtet. Somit gibt es in einer Evaluationsmatrix so viele MMMT-Matrizen wie verwendete Methoden. Jede MMMT-Matrix ist direkt an der Diagonalen mit Eigenkorrelationen zu finden. Die Eigenkorrelationen selbst werden von den Berechnungen aufgrund ihrer Trivialität ausgenommen. Bei der Analyse der MMMT-Matrix ist davon auszugehen, dass die verwendeten Mess-Items das Qualitätskriterium selbst ausreichend beschreiben, d.h. dass zwar bedeutende Korrelatio-

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

nen vorhanden sind, diese jedoch nicht zu hoch sind, so dass die Verwendung von verschiedenen Mess-Items für die jeweilige Methode gerechtfertigt ist.

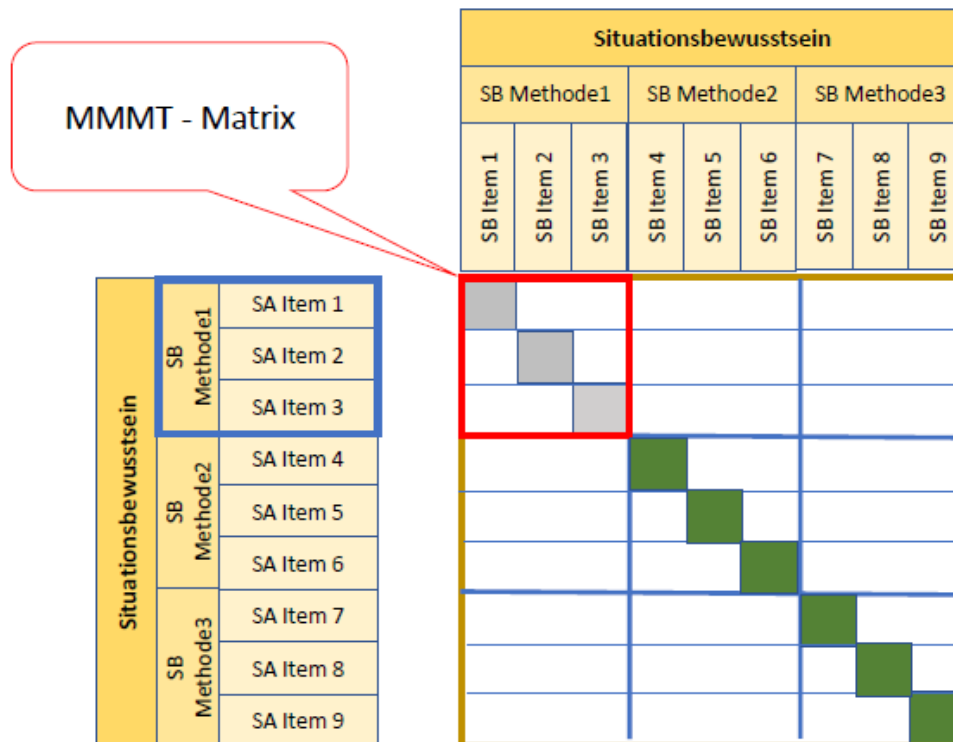


Abbildung 6.4 Schematische Illustration einer MMT-Matrix für das Qualitätskriterium Situationsbewusstsein

Im weiteren Verlauf werden die MMT-Matrizen zwar nominell der konvergenten Validität zugeordnet, da sie sich auf das gleiche Konstrukt wie die HMMT-Matrizen beziehen, sie werden jedoch bei der Auswertung jeweils separat betrachtet, da hier ein geringeres Maß an Korrelationsstärke erwünscht ist, als in den HMMT-Matrizen.

Die HMMT-Matrizen hingegen beinhalten den eigentlich zentralen Aspekt der konvergenten Validität, da hier strukturiert erfasst wird, wie stark eine Methode mit allen anderen Methoden eines einzigen Qualitätskriteriums korreliert. Hier wird erwartet, dass der Anteil an bedeutsamen Korrelationen besonders hoch ist, da jede Methode für sich in Anspruch nimmt, das Konstrukt oder zumindest deren Aspekte ausreichend zu beschreiben. Auch für die HMMT-Matrizen gilt, dass es so viele HMMT-Matrizen gibt wie eingesetzte Methoden. Analysiert wird jedoch der an diese Methode angrenzende Bereich innerhalb eines Qualitätskriteriums (vgl. Abbildung 6.5).

Vergleicht man nun die MMT-Matrizen mit den HMMT-Matrizen ist davon auszugehen, dass der Anteil bedeutsamer Korrelationen innerhalb der HMMT-Matrizen höher ist, als der Anteil bei den MMT-Matrizen, da bei den MMT-Matrizen die Items ausreichend trennscharf sein sollten und die Korrelationen der HMMT Konstrukte das gleiche Konstrukt erfassen.

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

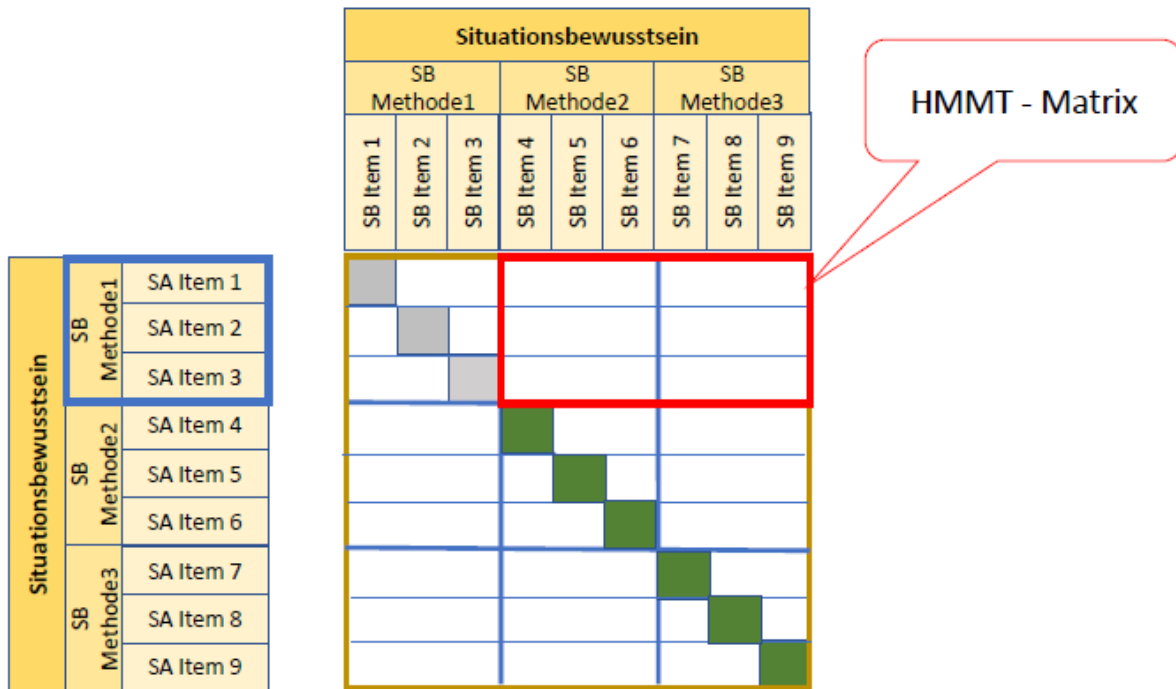


Abbildung 6.5 Schematische Illustration einer HMHT-Matrix für das Qualitätskriterium Situationsbewusstsein

6.2.2 Sub-Matrizen zur Bestimmung diskriminanter Validität

Für die Erfassung der diskriminanten Validität wird die Evaluationsmatrix aufgrund der gewählten Qualitätskriterien erneut in zwei Sub-Matrizen zerlegt, die jedoch unter dem Begriff der HeteroMethod-HeteroTrait-Matrix (HMHT-Matrix) zusammengefasst werden. Betrachtet werden die Korrelationen einer Methode mit den jeweils anderen Qualitätskriterien (vgl. Abbildung 6.6). So wird beispielsweise eine Methode des Situationsbewusstseins anhand der Qualitätskriterien Workload und Gebrauchstauglichkeit beschrieben. Die Submatrizen werden hierzu jeweils horizontal, d.h. von links nach rechts, gelesen. Das führt dazu, dass beispielsweise eine HMHT-Matrix, die auf einer Kombination von Situationsbewusstsein mit Workload beruht, nicht identisch mit einer HMHT-Matrix ist, bei der eine Workload-Methode mit dem Konstrukt Situationsbewusstsein korreliert wurde. Das heißt, es werden stets Korrelationen unterschiedlicher Mess-Items betrachtet.

Es ergeben sich somit folgende Matrizen-Kombinationen für die HMHT-Matrizen:

- Situationsbewusstsein mit Workload SB-WL
- Situationsbewusstsein mit Gebrauchstauglichkeit SB-GT
- Workload mit Situationsbewusstsein WL-SB
- Workload mit Gebrauchstauglichkeit WL-GT
- Gebrauchstauglichkeit mit Situationsbewusstsein GT-SB
- Gebrauchstauglichkeit mit Workload GT-WL

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

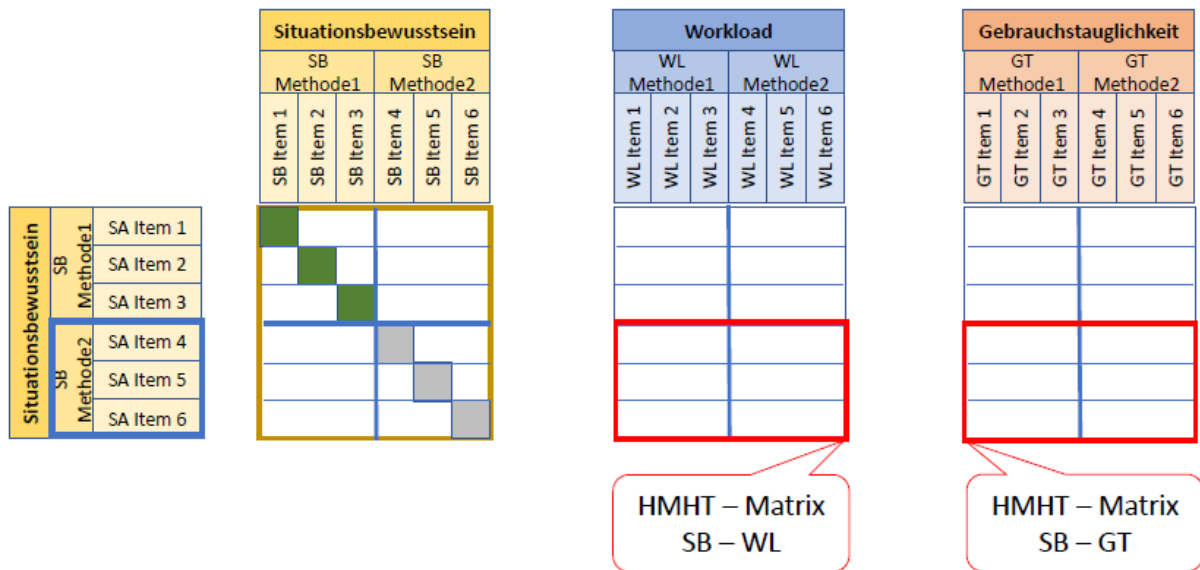


Abbildung 6.6 Schematische Illustration der HMHT-Matrizen für die Kombinationen Situationsbewusstsein mit Workload (SB-WL) und Situationsbewusstsein mit Gebrauchstauglichkeit (SB-GT)

Da mit Hilfe dieses Vorgehens die diskriminante Validität bestimmt wird, wird erwartet, dass die vorhandenen Korrelationen möglichst gering sind, da sich die Qualitätskriterien und ihre zu Grunde liegenden Konstrukte ausreichend voneinander unterscheiden sollten.

6.2.3 Interpretation der Analysematrizen

Jede Zelle der Evaluationsmatrix besteht, wie beschrieben, aus Item-Korrelationen. Jedes Item wird mit jedem anderen Mess-Item korreliert, was zu einer Spiegelung der Matrix in der Diagonalen führt. Da die korrelierten Items sowohl intervall- wie auch rangskaliert sind und Experten-Stichproben mit kleinem N in der Evaluierung verwendet wurden, wird der Spearman Korrelationskoeffizient ρ für die Validierung genutzt. Für die weitere Analyse der Matrix wird jeweils der prozentuale Anteil an unbedeutenden ($|\rho| \approx 0$), schwachen ($0.2 < |\rho| \leq 0.4$) und bedeutsamen Korrelationen ($|\rho| > 0.4$) in den Sub-Matrizen als Betrag und somit ohne Tendenzen erfasst. Die Sub-Matrizen lassen sich dann sowohl global – auf Konstrukt-Ebene sowie auf Methoden-Ebene oder auch auf Item-Ebene – interpretieren.

Zur Beschreibung der Qualitätskriterien werden die Methoden in den Evaluationsmatrizen für jedes Konstrukt zusammengefasst. Das heißt, alle Korrelationen der verwendeten Methoden eines Konstruktes werden gebündelt betrachtet und auch gebündelt mit den anderen Konstrukten verglichen. Beispielsweise werden zur Bewertung von Situationsbewusstsein alle MMT-Matrizen der Situationsbewusstseinsmethoden gemeinsam und auch alle HMHT-Matrizen der Situationsbewusstseinsmethoden betrachtet, um die konvergente Validität zu beschreiben. Zur Erhebung der diskriminanten Validität werden die Korrelationsstärken aller in einem Experiment verwendeten Situationsbewusstseinsmethoden mit den Konstrukten Workload und Gebrauchstauglichkeit prozentual erfasst.

6 Der Multi-Trait/Multi-Method-Ansatz als Strategie zur Erhebung der Konstrukt-Validität

Um zu bestimmen, wie gut die verwendeten Methoden im Einzelnen das jeweilige Konstrukt bzw. das Qualitätskriterium beschreiben, werden die Mess-Items zu Methoden zusammengefasst und es wird überprüft, wie gut sich diese zur Bestimmung der Qualitätskriterien in den einzelnen Evaluierungsexperimenten eignen, in denen sie verwendet wurden. Das heißt, die Methoden-Ebene ist eine detailliertere Sichtweise auf die Evaluationsmatrizen. Hierbei werden für jedes Experiment nur die Zeilen der Mess-Items einer Methode und die so entstehenden Sub-Matrizen betrachtet.

7 ERGEBNISSE DER VALIDIERUNG AUF KONSTRUKT- UND METHODEN-EBENE

Als Grundlage für die Validierung des multidimensionalen Methoden-Inventars werden die Korrelationsmatrizen der Evaluierungen (Kapitel 5) genutzt. Die Analyse der Matrizen erfolgt auf Konstrukt- und auf Methoden-Ebene sowie unter Betrachtung der jeweiligen konvergenten und diskriminanten Validität. Zur Betrachtung der Konstrukt-Ebene bzw. der Qualitätskriterien (s.u. Kapitel 7.1) werden die Methoden für jedes Qualitätskriterium innerhalb der durchgeführten Studien zusammengefasst und zur Beschreibung der konvergenten und der diskriminanten Validitäten genutzt. Die Beschreibung der Validierungsergebnisse auf Methoden-Ebene erfolgt hingegen für jede Methode separat und wird anhand der jeweiligen Evaluierungen vorgenommen, in denen die Methoden verwendet wurden (vgl. Kapitel 7.2).

7.1 Validierungsergebnisse der Qualitätskriterien auf Konstrukt-Ebene

Für die Evaluierungen, die in Kapitel 5 vorgestellt wurden, sind die Qualitätskriterien zur Bewertung von technischen Funktionen und deren Bedienung mit verschiedenen Methoden des multidimensionalen Methoden-Inventars erfasst worden. Bei näherer Betrachtung der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit ist daher zu erwarten, dass die Methoden, die ein und dasselbe Qualitätskriterium messen, höher miteinander korrelieren als mit den anderen beiden Qualitätskriterien. Die konvergente Validität wird daher über alle MMT- und HMMT-Matrizen für alle vier Evaluationen für jedes Qualitätskriterium betrachtet und mit der diskriminanten Validität, die anhand der HMHT-Matrizen beschrieben wird, verglichen. Die Methoden eines Qualitätskriteriums werden somit zusammengefasst und als Ganzes bezüglich ihrer Messgenauigkeit für das jeweilige Konstrukt bewertet.

7.1.1 Situationsbewusstsein

Die Analyse der MMT-Matrizen (vgl. Tabelle 7.1 und Abbildung 7.1), welche die Korrelationen innerhalb der jeweiligen Methoden zur Erfassung des Situationsbewusstseins beschreiben, zeigt, dass die Messwerte stark miteinander korrelieren. Das ist insbesondere bei der Machbarkeitsstudie der Navigationshilfe (MS) mit 91.67 % an bedeutsamen Korrelationen auffällig. Beachtenswert sind auch die Studien zum Re-Design der Navigationshilfe (RD) sowie die Untersuchung des räumlich-auditiven Warnsignals (3DA), in welchen das Methoden-Inventar zur Erfassung des Situationsbewusstseins umfangreicher genutzt wurde. Auch hier sind die Korrelationen innerhalb der angewendeten Methoden relativ hoch. Das gilt insbesondere für die Untersuchung des Re-Designs mit 54.20 % an bedeutsamen Korrelationen innerhalb der Methoden. Deutlich weniger hoch ist der Prozentsatz bei der Untersuchung des räumlich-auditiven Warnsignals mit 38.04 %. Auch in der letzten Studie zur Untersuchung der Genauigkeit und Sensitivität eines Sensors (SEN) ist der Prozentsatz an bedeutsamen Korrelationen innerhalb der Methoden, die zur Erfassung des Situationsbewusstseins genutzt wurden, mit 68.18 % eher hoch.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.1 Prozentualer Anteil an Korrelationsstärken in den MMMT-Matrizen zur Erfassung des Situationsbewusstseins

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	23.35	28.15	20.10
$0.2 < \rho \leq 0.4$	8.33	22.44	33.81	11.72
$ \rho > 0.4$	91.67	54.20	38.04	68.18

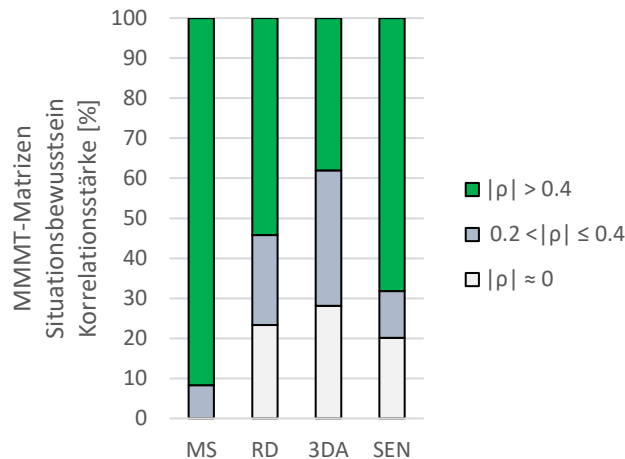


Abbildung 7.1 Prozentualer Anteil an Korrelationsstärken in den MMMT-Matrizen zur Erfassung des Situationsbewusstseins

Der Kern der konvergenten Validität sind die HMMT-Matrizen. Jede einzelne Methode, die zur Erfassung des Situationsbewusstseins verwendet wurde, wird hier mit den anderen Methoden, die ebenso zur Erfassung des Situationsbewusstseins verwendet wurden, korreliert und über alle Evaluierungen in Tabelle 7.2 und Abbildung 7.2 verglichen. In der Machbarkeitsstudie wurden zwei Herangehensweisen des Eye-Trackings miteinander verglichen und erreichen 100 % an bedeutsamen Korrelationen. Bei einer Korrelation des 3D SARTs mit Eye-Tracking im Rahmen der Sensorstudie ergeben sich hingegen 39.19 % an bedeutsamen Korrelationen. Die bedeutsamen Korrelationen der Evaluierungen der Re-Design-Studie (10.98 %) und der Untersuchung des räumlich-auditiven Warnsignals (7.67 %) zeigen hingegen geringere Werte bezüglich der konvergenten Validität auf. So liegt der Anteil an Nichtkorrelationen bei RD bei 57.49 % und innerhalb der 3DA Evaluierung bei 54.31 %.

Tabelle 7.2 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Erfassung des Situationsbewusstseins

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	57.49	54.31	31.43
$0.2 < \rho \leq 0.4$	0.00	31.53	38.02	29.38
$ \rho > 0.4$	100.00	10.98	7.67	39.19

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

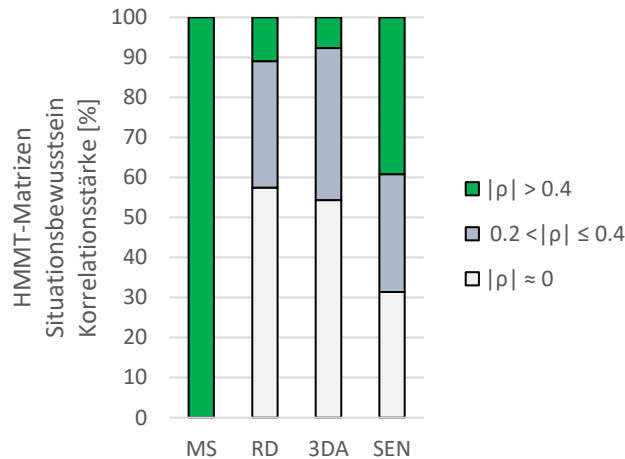


Abbildung 7.2 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Erfassung des Situationsbewusstseins

Dieser Eindruck wird bei Analyse der diskriminanten Validität anhand der HMHT-Matrizen unter Betrachtung der Korrelationen der Methoden zur Erfassung des Situationsbewusstseins in Kombination mit den Konstrukten Workload (vgl. Tabelle 7.3 und Abbildung 7.3) und Gebrauchstauglichkeit (vgl. Tabelle 7.4 und Abbildung 7.4) bestätigt.

Tabelle 7.3 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Workload

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	41.96	53.77	50.47	29.81
$0.2 < \rho \leq 0.4$	22.32	30.72	33.77	24.55
$ \rho > 0.4$	35.71	15.51	15.76	45.64

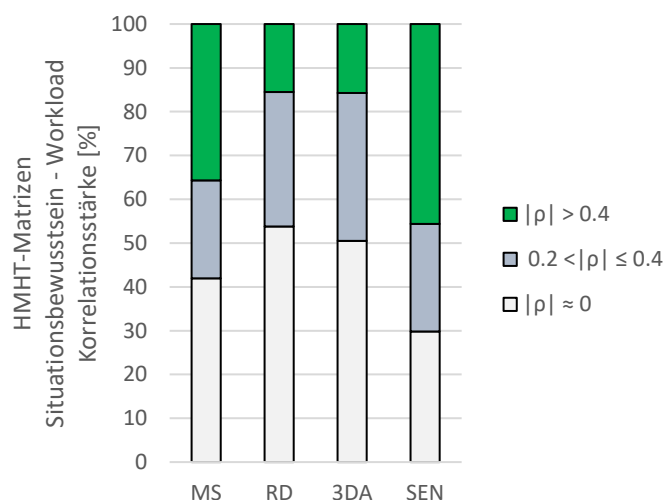


Abbildung 7.3 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Workload

Deutlich wird, dass mit Ausnahme der Machbarkeitsstudie die mit 35.71 % in der Kombination Situationsbewusstsein-Workload (SB-WL) und mit 30.51 % in der Situationsbewusstsein-Gebrauchstauglichkeit-Kombination (SB-GB) die HMHT-Matrizen einen höheren Anteil an bedeutsa-

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

men Korrelationen aufweisen als im Vergleich hierzu die HMHT-Matrizen. Dies gilt für die verbleibenden Evaluationen sowohl für die SB-WL-Kombination als auch für die SB-GT-Kombination. Sowohl die 3DA-Studie (15.76 %) als auch die SEN-Studie (45.64 %) weisen in der SB-WL-Kombination einen höheren Anteil an bedeutsamen Korrelationen auf als bei Betrachtung der HMHT-Matrizen. Das gilt auch für das Re-Design der Navigationshilfe mit 15.51 %. Allerdings scheinen in der RD-Evaluation die Methoden zur Erfassung des Situationsbewusstseins marginal stärker mit den Methoden der Gebrauchstauglichkeit zu korrelieren – bzw. dieses Konstrukt besser zu erklären, da hier der Anteil an bedeutsamen Korrelationen bei 17 % liegt. Auch für die anderen beiden Evaluierungen zeigt sich, dass hier die Methoden zur Erfassung des Situationsbewusstseins mit denen der Gebrauchstauglichkeit stärker korrelieren als im Vergleich zu den HMHT-Matrizen. Der Anteil der bedeutsamen Korrelationen ist bei den verbleibenden Evaluierungen in der SB-GT-Kombination (3DA = 13.65 %, SEN = 39.67 %) jedoch leicht niedriger als in der SB-WL-Kombination. Hinzu kommt, dass auch der Anteil ohne relevante Zusammenhänge in den HMHT-Matrizen ähnlich hoch ist wie in den HMHT-Matrizen.

Tabelle 7.4 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Gebrauchstauglichkeit

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	45.71	49.83	50.59	32.96
$0.2 < \rho \leq 0.4$	23.77	33.18	35.76	27.37
$ \rho > 0.4$	30.51	17.00	13.65	39.67

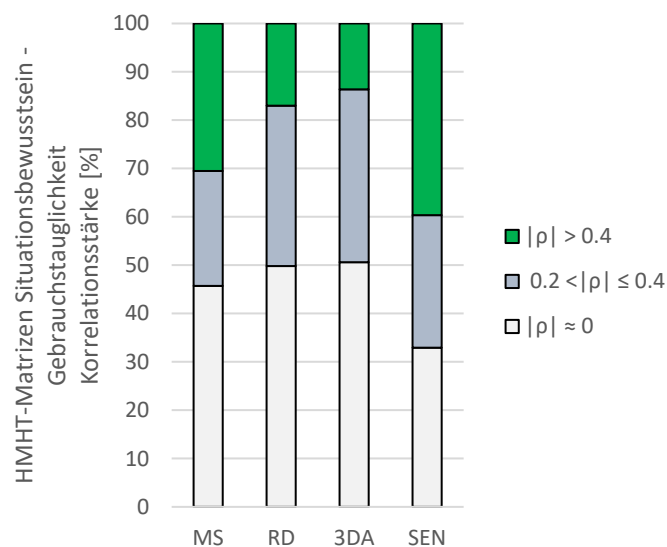


Abbildung 7.4 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Gebrauchstauglichkeit

7.1.2 Workload

Auch bei der Analyse des Konstruktes Workload über alle Evaluierungen hinweg zeigt sich, dass die MMT-Matrizen (vgl. Tabelle 7.5 und Abbildung 7.5) einen hohen Anteil an bedeutsamen Korrelationen aufweisen. Der Anteil an bedeutenden Korrelationen in den MMT-Matrizen ist für alle Evaluierungen größer als 50 %. Der Anteil an vernachlässigbaren Korrelationen liegt hierbei bei allen Stu-

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

dien bei unter 15 % – mit Ausnahme der Evaluierung der Sensoreinstellungen mit 27.38 %. Der Anteil an Nichtkorrelationen schwankt hingegen stark mit einem sehr geringen Anteil in der RD-Evaluierung (1.90 %) und einen vergleichsweise hohen Anteil in der Untersuchung des räumlich-auditiven Warnsignals mit 35.24 %.

Tabelle 7.5 Prozentualer Anteil an Korrelationsstärken in den MMMT-Matrizen zur Beschreibung der Workload

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	26.19	1.90	35.24	9.52
$0.2 < \rho \leq 0.4$	1.19	13.71	10.95	27.38
$ \rho > 0.4$	72.62	84.38	53.81	63.10

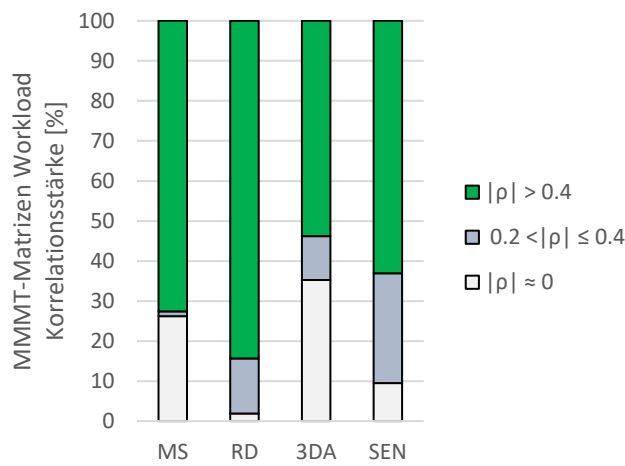


Abbildung 7.5 Prozentualer Anteil an Korrelationsstärken in den MMMT-Matrizen zur Beschreibung der Workload

Bei Analyse der HMMT-Matrizen (vgl. Tabelle 7.6 und Abbildung 7.6) wird hingegen erneut ein eher geringer Anteil an bedeutsamen Korrelationen offensichtlich. Zwar weisen die Machbarkeitsstudie mit 41.85 % und die Evaluierung der Sensoreinstellung mit 38.73 % an bedeutsamen Korrelationen auf eine nicht zu vernachlässigende konvergente Validität hin, jedoch relativiert sich dieser Eindruck nach Berücksichtigung der anderen beiden Untersuchungen (RD = 8.13 %, 3DA = 8.57 %).

Tabelle 7.6 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Workload

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	22.85	59.82	64.88	43.00
$0.2 < \rho \leq 0.4$	35.30	32.05	26.55	18.27
$ \rho > 0.4$	41.85	8.13	8.57	38.73

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

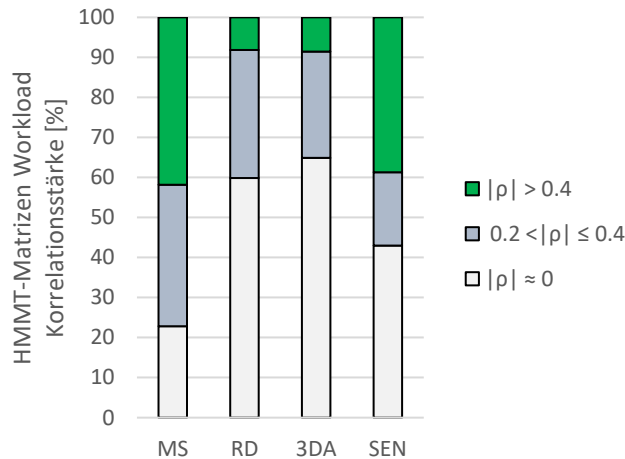


Abbildung 7.6 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Workload

Im Vergleich zur konvergenten Validität ist die diskriminante Validität in der WL-SB-Kombination, mit erneuter Ausnahme der Machbarkeitsstudie mit 35.20 %, immer höher. Vor allem stellt sich aber heraus, dass auch der Anteil an Nichtkorrelationen insbesondere in den HMMT-Matrizen sehr hoch ist. Auffällig ist, dass bei umfänglicher Anwendung des Methoden-Inventars in der Re-Design-Studie sowie bei Evaluierung des räumlich-auditiven Warnsignals der Anteil an bedeutsamen Zusammenhängen deutlich niedriger ist als in den anderen beiden kleineren Evaluierungen. Dies gilt sowohl für die HMMT-Matrizen als auch für die HMHT-Matrizen. So liegt der Anteil an bedeutsamen Korrelationen in der HMMT-Matrix der RD-Evaluierung bei 8.13 % und in der HMMT-Matrix der 3D-Evaluierung bei 8.57 %. In den HMHT-Matrizen für die WL-SB-Kombination (RD = 15.85 %, 3DA = 12.81 %) und für die WL-GT-Kombination (RD = 21.30 %, 3DA = 12.69 %) sind die Werte jedoch höher. Auch die Sensorstudie zeigt in der HMHT-Matrix der SB-WL-Kombination einen höheren Anteil an bedeutsamen Korrelationen (46.57 %) als hinsichtlich der HMMT-Matrix (38.73 %).

Tabelle 7.7 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	46.68	53.06	55.44	29.82
$0.2 < \rho \leq 0.4$	18.11	31.08	31.74	23.61
$ \rho > 0.4$	35.20	15.85	12.81	46.57

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

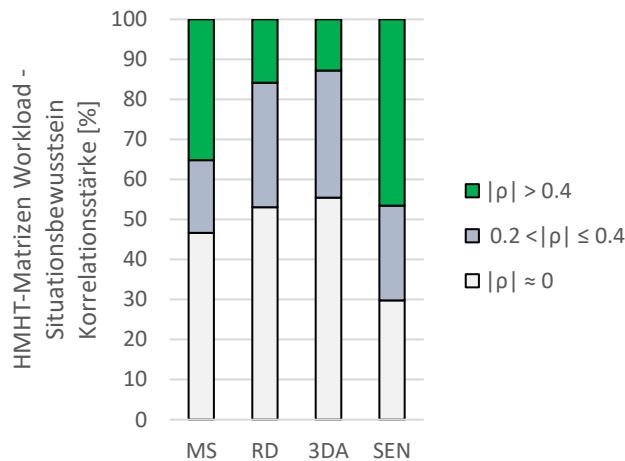


Abbildung 7.7 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Situationsbewusstsein

Die Gegenüberstellung der HMHT-Matrizen der Workload-Methoden mit den Matrizen der WL-GT-Kombinationen (vgl. Tabelle 7.8 und Abbildung 7.8) zeigt, dass alle Werte für bedeutsame Korrelationen in den HMHT-Matrizen höher sind als vergleichsweise innerhalb HMHT-Matrizen. Die 3DA-Studie weist mit 12.81 % in der WL-SB-Kombination und mit 12.69 % in der WL-GT-Kombination einen ähnlichen Anteil an diskriminanter Validität in beiden Kombinationen auf. Die Sensorstudie hingegen zeigt einen niedrigeren Anteil in der WL-GT-Kombination (40.29 %) auf als in der WL-SB-Kombination (46.57 %). In der RD-Evaluierung ist der Anteil an bedeutsamen Korrelationen hingegen in der WL-SB-Kombination (15.85 %) niedriger als in der WL-GT-Kombination (21.30 %). Auch hier bildet die Machbarkeitsstudie eine Ausnahme, da die diskriminante Validität in der WL-SB-Kombination niedriger ist (35.20 %) als die konvergente Validität in der HMHT-Matrix (41.85 %) jedoch eine höhere Ausprägung an bedeutsamen Korrelationen in der WL-GT-Kombination aufweist (45.48 %).

Tabelle 7.8 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Gebrauchstauglichkeit

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	26.52	43.57	54.70	34.57
$0.2 < \rho \leq 0.4$	27.99	35.13	32.61	25.14
$ \rho > 0.4$	45.48	21.30	12.69	40.29

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

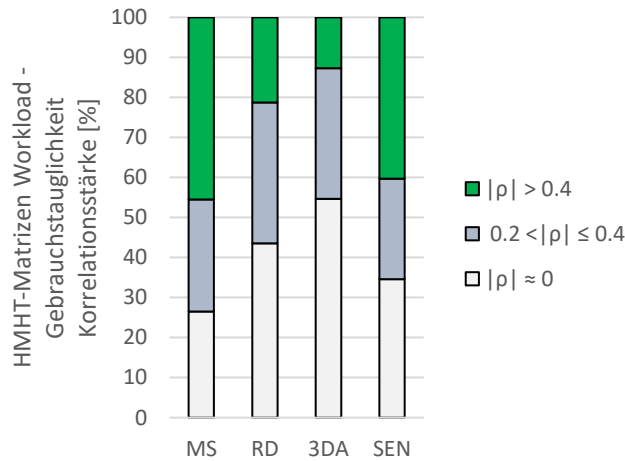


Abbildung 7.8 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Gebrauchstauglichkeit

7.1.3 Gebrauchstauglichkeit

Die Mess-Items zur Bestimmung der Gebrauchstauglichkeit deuten auf einen niedrigeren Anteil an bedeutsamen Korrelationen in den MMT-Matrizen hin. In der Machbarkeitsstudie ist der Anteil an bedeutsamen Korrelationen mit 73.26 % dennoch vergleichsweise hoch. Die Studien RD mit 41.71 %, 3DA mit 50.59 % und SEN mit 32.64 % hingegen zeigen deutlich geringere Ausprägungen an bedeutsamen Korrelationen auf und deuten auf eine effiziente Konstellation der Mess-Items hin. Die weiteren Ergebnisse der MMT-Matrizen sind in Tabelle 7.9 und Abbildung 7.9 zu finden.

Tabelle 7.9 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	8.00	35.37	30.16	21.53
$0.2 < \rho \leq 0.4$	18.74	22.92	18.89	45.83
$ \rho > 0.4$	73.26	41.71	50.95	32.64

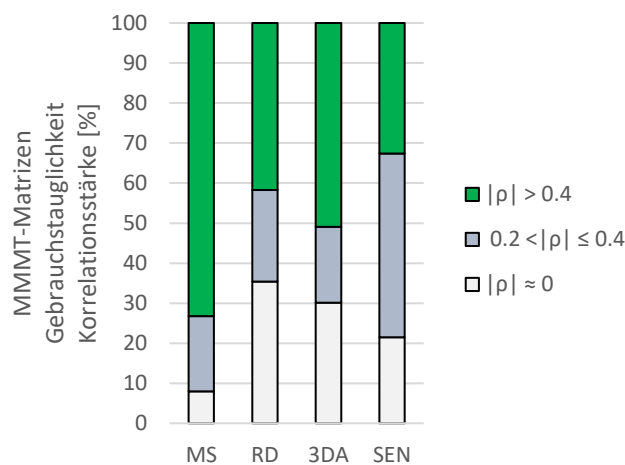


Abbildung 7.9 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Mit Blick auf die HMMT-Matrizen (vgl. Tabelle 7.10 und Abbildung 7.10) zeigt sich ein ambivalentes Bild im Hinblick auf die konvergente Validität. In der Machbarkeitsstudie ist der Anteil an bedeutsamen Korrelationen mit 34.51 % niedriger und auch in der Evaluation der Sensorbedienung zeigt sich mit 32.51 % ein kleinerer Anteil als im Vergleich mit den anderen beiden Konstrukten. Das Gegenteil wird jedoch deutlich bei der Analyse der 3DA-Studie (12.52 %) und der Evaluierung des Re-Designs (23.38 %), da diese beiden Evaluierungen einen höheren Anteil an bedeutsamen Korrelationen und somit konvergenter Validität aufweisen als die Qualitätskriterien Situationsbewusstsein oder Workload.

Tabelle 7.10 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	36.06	37.05	54.89	32.04
$0.2 < \rho \leq 0.4$	29.43	39.58	32.58	35.45
$ \rho > 0.4$	34.51	23.38	12.52	32.51

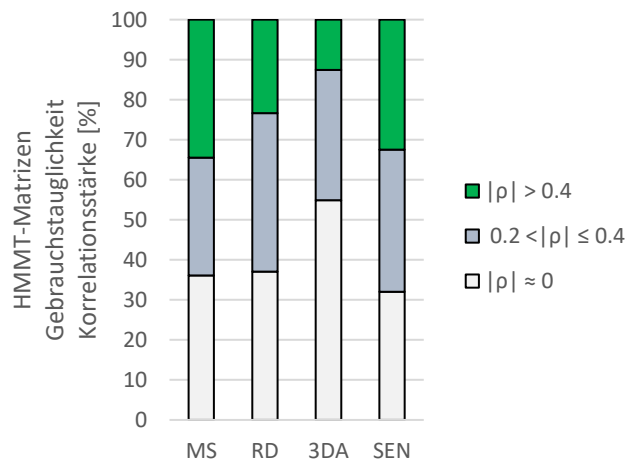


Abbildung 7.10 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit

Bei der Gegenüberstellung der HMMT-Matrizen mit den HMHT-Matrizen zeigt sich des Weiteren sowohl für die GT-SB-Kombination (vgl. Tabelle 7.11 und Abbildung 7.11) als auch für die GT-WL-Kombination (vgl. Tabelle 7.12 und Abbildung 7.12), dass die Anteile an bedeutsamen Korrelationen für die diskriminante Validität in diesen beiden Studien geringer sind als die Werte in den HMMT-Matrizen. So sind sowohl die Anteile bedeutsamer Korrelationen in der GT-SB-Kombination (RD = 20.89 %, 3DA = 9.79 %) als auch in der GT-WL-Kombination geringer (RD = 22.84 %, 3DA = 10.96 %) als die Werte in den HMMT-Matrizen (RD = 23.38 %, 3DA = 12,52 %). Betrachtet man hingegen die Machbarkeitsstudie und die SEN-Evaluierung, ist das Gegenteil der Fall, da hier Anteile der diskriminanten Validität höher sind als die der konvergenten Validität. Hier zeigen sich sowohl in der GT-SB-Kombination (MS = 38.52 %, SEN = 41.47 %) als auch in der GT-WL-Kombination (MS = 58.76 %, SEN = 35.33 %) höhere Anteile an bedeutsamen Korrelationen als in den HMMT-Matrizen (MS = 34.51 %, SEN = 32.51 %).

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.11 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Situationsbewusstsein

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	36.52	49.35	56.22	29.22
$0.2 < \rho \leq 0.4$	24.95	29.76	33.99	29.31
$ \rho > 0.4$	38.52	20.89	9.79	41.47

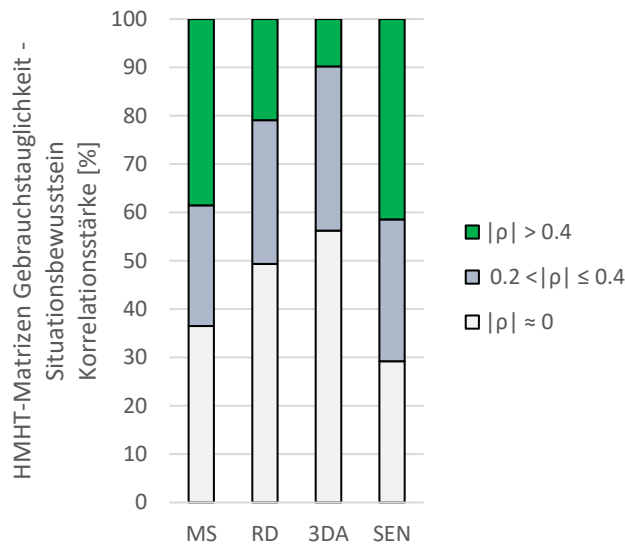


Abbildung 7.11 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Situationsbewusstsein

Bei Betrachtung der beiden HMHT-Matrizen und den jeweiligen Kombinationen scheint Gebrauchstauglichkeit etwas mehr mit Workload zusammenzuhängen (RD = 22.84 %, 3DA = 10.96 %, MS = 58.76 %) als mit Situationsbewusstsein (RD = 20.89 %, 3DA = 9.79 %, MS = 38.52 %). Lediglich in der SEN-Evaluierung wird Situationsbewusstsein (41.47 %) anhand der Methoden der Gebrauchstauglichkeit mit einem höheren Anteil an bedeutsamen Korrelationen beschrieben als Workload (35.33 %).

Tabelle 7.12 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Workload

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	21.83	43.13	54.64	35.90
$0.2 < \rho \leq 0.4$	19.40	34.04	34.40	28.77
$ \rho > 0.4$	58.76	22.84	10.96	35.33

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

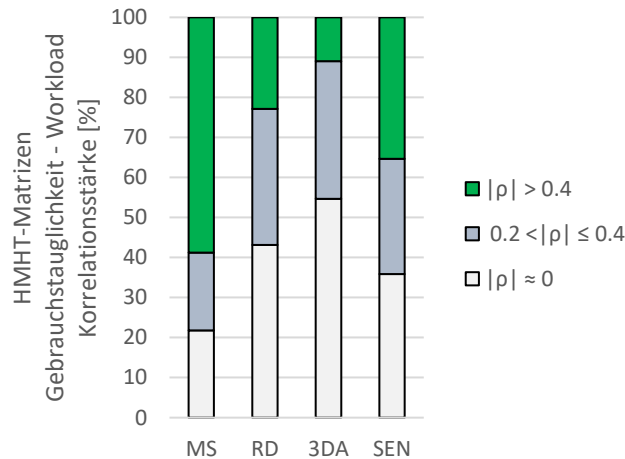


Abbildung 7.12 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Workload

7.1.4 Zusammenfassende Betrachtung der Konzepte

Werden die bedeutsamen Korrelationen für alle Evaluierungen für die MMT-Matrizen zusammengefasst, zeigt sich (vgl. Abbildung 7.13), dass das Konzept Gebrauchstauglichkeit (49.64 %) mit heterogeneren Items beschrieben wird als die Konstrukte Situationsbewusstsein (63.02 %) und Workload (68.48 %). Im Rückschluss wird daher deutlich, dass insbesondere die Qualitätskriterien Situationsbewusstsein und Workload hohe Anteile an bedeutsamen Korrelationen aufweisen. Demnach zeigt sich in der Gesamtschau der MMT-Matrizen, dass die verwendeten Mess-Items innerhalb der Methoden zur Erfassung des Situationsbewusstseins und der Workload stark miteinander korrelieren.

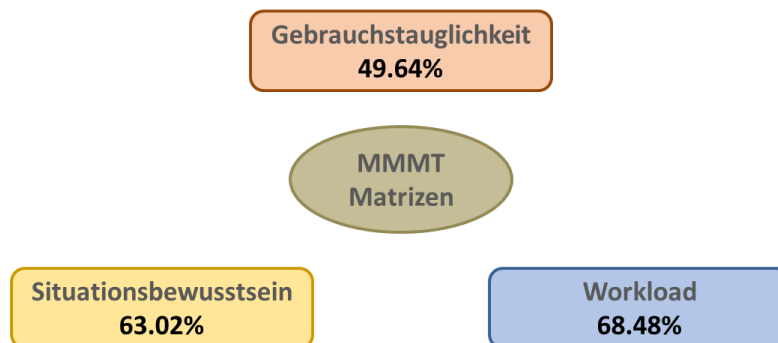


Abbildung 7.13 Anteil an bedeutsamen Korrelationen in den MMT-Matrizen über alle Evaluierungen

Werden die Evaluierungen zusammengefasst und konvergente und diskriminante Validität gegenübergestellt (vgl. Abbildung 7.14), wird deutlich, dass mit Ausnahme des Qualitätskriteriums Situationsbewusstsein und dessen Methoden (39.46 %) die diskriminante Validität höher als die konvergente Validität ist. Da dieser Wert jedoch die Matrizen aller Evaluierungen beinhaltet, muss festgehalten werden, dass die Machbarkeitsstudie diesen Wert verzerrt, da in dieser Studie zwei Eye-Tracking-Herangehensweisen gegenübergestellt wurden. Ignoriert man die Machbarkeitsstudie, läge der Anteil an bedeutsamen Korrelationen der HMHT-Matrizen für die Methoden zur Erfassung des Situationsbewusstseins lediglich bei 19.28 %. Somit wäre das Konzept Situationsbewusstsein durch die ei-

genen Methoden schlechter beschrieben als Workload mit 24.32 % und Gebrauchstauglichkeit mit 25.73 %.

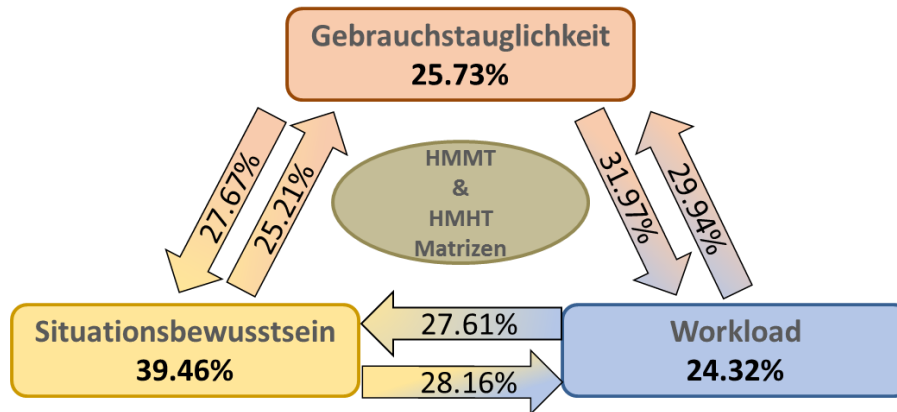


Abbildung 7.14 Gegenüberstellung der konvergenten und der diskriminanten Validität anhand des Anteils an bedeutsamen Korrelationen über alle Evaluierungen

Bei Analyse der HMHT-Matrizen zeigt sich, dass die verwendeten Methoden der Gebrauchstauglichkeit Workload mit 31.97 % am besten beschreiben. Auch im umgekehrten Fall ist mit 29.94 % eine deutliche Verbindung dieser beiden Konstrukte erkennbar. Die angewendeten Workload-Methoden messen Situationsbewusstseins mit 27.61 % bedeutsam, umgekehrt mit 28.61 %. Gebrauchstauglichkeit beschreibt das Situationsbewusstsein mit 27.67 % etwas besser als umgekehrt das Situationsbewusstsein die Gebrauchstauglichkeit mit 25.21 %.

Es wird insgesamt deutlich, dass die verwendeten Methoden und ihre Messwerte nicht zwischen den Konzepten diskriminieren. Wird bei Betrachtung der Methoden des Qualitätskriteriums Situationsbewusstsein zudem der korrigierte Wert angenommen, zeigt sich, dass konvergente und diskriminante Validität jeweils in einem ähnlichen Wertebereich liegen. Das Situationsbewusstsein ist dann bezüglich der konvergenten Validität sogar etwas niedriger zu bewerten, als die anderen beiden Konstrukte. Die diskriminante Validität betreffend, wird eine marginal stärkere Verknüpfung von Workload und Gebrauchstauglichkeit sichtbar. Das heißt, dass die Kombinationen WL-SB und SB-GT weniger stark miteinander zusammenhängen.

7.2 Validierungsergebnisse der Einzelmethode

Im Rahmen der Analyse der Konstrukt-Validität bezüglich der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit stellt sich heraus, dass die Konstrukte nicht disjunkt voneinander abzugrenzen sind, sondern im Gegenteil stark miteinander verbunden sind. Daher stellt sich die Frage, ob dies für alle verwendeten Methoden gilt oder ob einzelne Methoden besser dazu geeignet sind, die Qualitätskriterien zu beschreiben. Daher werden die einzelnen Methoden über alle Evaluierungsvorgehen hinsichtlich ihrer konvergenten und diskriminanten Validität überprüft. In die Validierung sind hierbei nur die Methoden eingeflossen, denen ein quantitativer Datensatz zu Grun-

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

de liegt, d.h. jegliche Form von qualitativen Methoden (wie Interviews und der Cognitive Walkthrough) werden ausgeschlossen, da diese Methoden nicht in einen quantitativen Datensatz übertragen werden können. Die Methoden des Methoden-Inventars sind sukzessive in die Simulationsumgebung integriert worden und sind daher für die Aussagekraft der Validierung der verschiedenen Methoden von unterschiedlicher Bedeutung. Das heißt, dass Methoden, die in jeder Evaluierung angewendet worden sind, auf eine breitere Datenbasis zurückzuführen sind als Methoden, die nur in einer Studie Verwendung gefunden haben. Aus Tabelle 7.13 kann entnommen werden, welche Methodik in den jeweiligen Evaluationsstudien verwendet worden ist. Die Ergebnisse der Validierung werden für jede Methode und alle Konstrukte der Reihe nach in den Kapiteln für Situationsbewusstsein, Workload und Gebrauchstauglichkeit beschrieben.

Tabelle 7.13 Übersicht der validierten Methoden des Methoden-Inventars

	MS	RD	3DA	SEN
Situationsbewusstsein	Eye-Tracking	3D SART SAGAT Sortieraufgabe Eye-Tracking	3D SART SAGAT Gefahren-Detektierung Eye-Tracking	3D SART Eye-Tracking
Workload	NASA TLX Eye-Tracking	NASA TLX TDRT Eye-Tracking	NASA TLX TDRT Eye-Tracking	NASA TLX Eye-Tracking
Gebrauchstauglichkeit	Nutzerleistung PSSUQ Eye-Tracking	Nutzerleistung PSSUQ Eye-Tracking	Nutzerleistung Eye-Tracking	Nutzerleistung Eye-Tracking

7.2.1 Methoden zur Erfassung des Situationsbewusstseins

Das Situationsbewusstsein ist mit verschiedenen subjektiven und objektiven Methoden und Messvorgehen strukturiert erfasst und sukzessive in die Evaluierungen integriert worden. Eine Analyse des Konzeptes Situationsbewusstsein in Kapitel 7.1.1 und Kapitel 7.1.4 ergab, dass nur unter Beibehaltung der Machbarkeitsstudie eine höhere konvergente Validität besteht. Werden die Ergebnisse der Machbarkeitsstudie ignoriert, wird das Konstrukt Situationsbewusstsein durch die Methoden der Workload und der Gebrauchstauglichkeit besser beschrieben. Zur genaueren Analyse dieses Ergebnisses werden die einzelnen Methoden, die zur Erfassung des Situationsbewusstseins genutzt werden, gesondert analysiert.

7.2.1.1 3D Situation Awareness Rating Technique (3D SART)

Bei Betrachtung der MMT-Matrizen des 3D SARTs (vgl. Tabelle 7.14 und Abbildung 7.15) zeigen sich hohe Anteile an bedeutsamen Korrelationen – insbesondere aber in der Studie zur Untersuchung der Sensoreinstellungen mit 83.33 %. In den breiter angelegten Studien zur Untersuchung des Re-Designs und des räumlich-auditiven Warnsignals ist dieser Anteil mit jeweils 50 % deutlich geringer.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.14 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des 3D SARTs

Spearman's ρ in %	RD	3DA	SEN
$ \rho \approx 0$	33.33	16.67	16.67
$0.2 < \rho \leq 0.4$	16.67	33.33	0.00
$ \rho > 0.4$	50.00	50.00	83.33

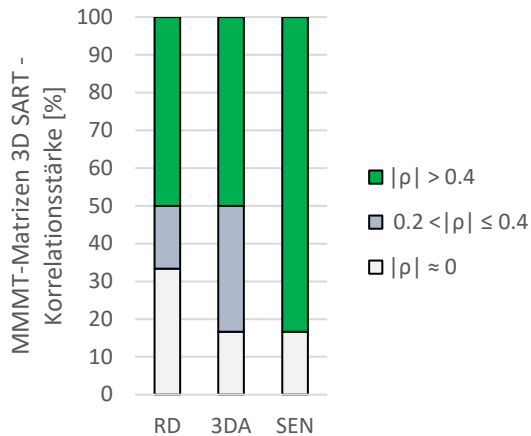


Abbildung 7.15 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des 3D SARTs

Tabelle 7.15 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des 3D SARTs

	RD	3DA	SEN
$ \rho \approx 0$	54.17	54.17	35.29
$0.2 < \rho \leq 0.4$	35.83	34.38	25.00
$ \rho > 0.4$	10.00	11.46	39.71

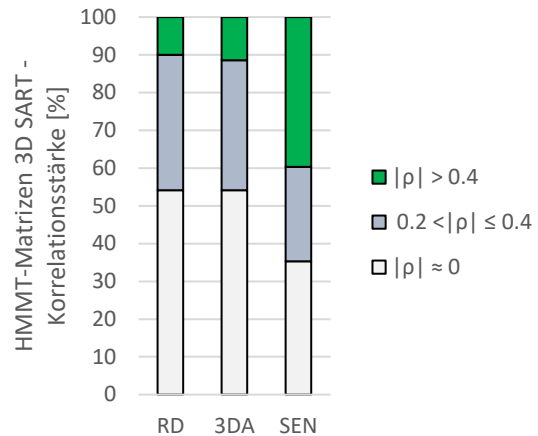


Abbildung 7.16 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des 3D SARTs

Stellt man die konvergente Validität (vgl. Tabelle 7.15 und Abbildung 7.16) der diskriminanten Validität (vgl. Tabelle 7.16 und Abbildung 7.17 sowie Tabelle 7.17 und Abbildung 7.18) gegenüber, zeigt sich ein ambivalentes Bild. In der RD-Untersuchung ist die konvergente Validität mit 10 % an bedeutsamen Korrelationen deutlich niedriger als die diskriminante Validität des 3D SARTs in Kombination mit Workload (RD = 19.57 %) und etwas niedriger als in der Kombination mit den Methoden der Gebrauchstauglichkeit (RD = 12.78 %). Umgekehrt verhält es sich allerdings bei der Analyse der 3DA Evaluation. Hier ist die diskriminante Validität sowohl in der Workload-Kombination (3DA = 10.23 %) als auch in der Gebrauchstauglichkeitskombination (3DA = 7.64 %) niedriger als die konvergente Validität mit 11.46 %. Allerdings sollte hier ebenso der hohe Anteil an Nichtkorrelationen mit 54.17 % beachtet werden, der sich auch in der RD-Studie mit 54.17 % zeigt. Die Ergebnisse der Sensorstudie sind hingegen schwerer zu beschreiben, da hier der 3D SART mit Hilfe der Workload besser aufgeklärt wird (SEN = 56.94 %) als durch die verwendeten Methoden zur Erfassung des Situationsbewusstseins (SEN = 39.71 %), die HMMT-Matrix jedoch einen höheren Wert als die 3DSART-Gebrauchstauglichkeitskombination (SEN = 30.77 %) aufweist.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.16 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Workload

Spearman's ρ in %	RD	3DA	SEN
$ \rho \approx 0$	51.09	61.36	20.83
$0.2 < \rho \leq 0.4$	29.35	28.41	22.22
$ \rho > 0.4$	19.57	10.23	56.94

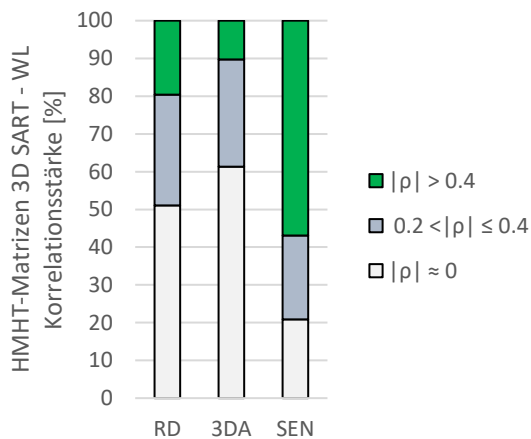


Abbildung 7.17 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Workload

Tabelle 7.17 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Gebrauchstauglichkeit

	RD	3DA	SEN
	53.33	56.94	40.38
	33.89	35.42	28,85
	12.78	7.64	30.77

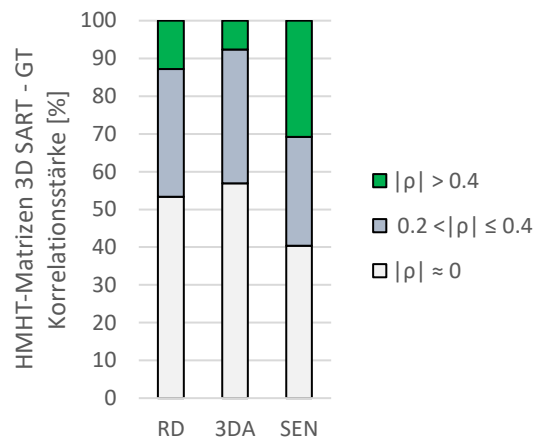


Abbildung 7.18 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Gebrauchstauglichkeit

7.2.1.2 Situation Awareness Global Assessment Technique (SAGAT)

Der SAGAT ist in den Erhebungen zur Evaluierung des Re-Designs einer Navigationshilfe und zur Untersuchung des räumlich-auditiven Warnsignals verwendet worden. Beide Evaluierungen greifen auf eine etwas breitere Datenbasis und auf einen umfangreichen Einsatz des multidimensionalen Methoden-Inventars zurück. Die Analyse der MMMT-Matrizen weist darauf hin, dass die verwendeten Methodenitems innerhalb der jeweiligen Evaluierung verschieden stark zusammenhängen. Insbesondere die Items in der 3DA-Studie sind deutlich unabhängiger voneinander (3DA = 5.49 %) als die verwendeten Items in der RD-Untersuchung mit 25.45 %.

Tabelle 7.18 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des SAGATs

Spearman's ρ in %	RD	3DA
$ \rho \approx 0$	34.55	65.93
$0.2 < \rho \leq 0.4$	40.00	28.57
$ \rho > 0.4$	25.45	5.49

Tabelle 7.19 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs

	RD	3DA
	60.08	58.16
	33.20	35.20
	6.72	6.63

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

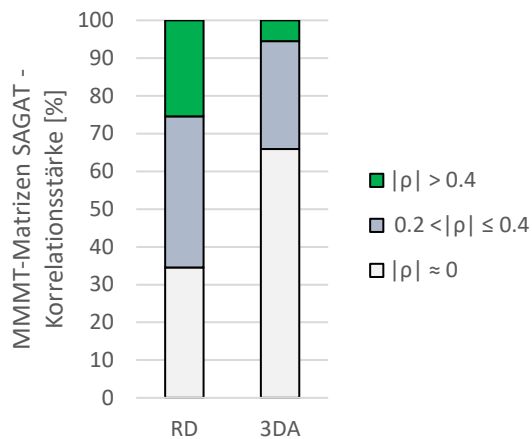


Abbildung 7.19 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des SAGATs

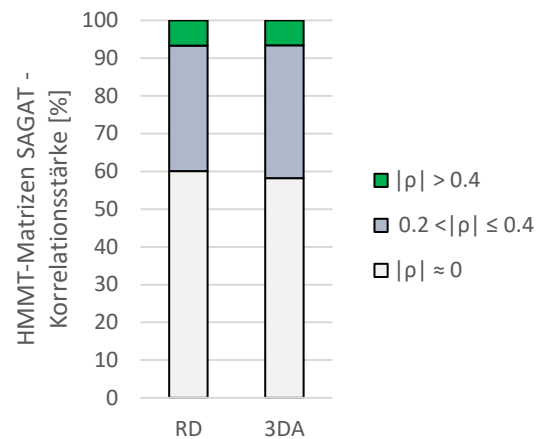


Abbildung 7.20 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs

Bei Betrachtung der HMMT-Matrizen ist vor allem die niedrige Aufklärung in beiden Studien durch Verwendung des SAGATs als Methodik zur Beschreibung des Situationsbewusstseins augenscheinlich (RD = 6.72 %, 3DA = 6.63 %). Die Aufklärung der SAGAT-Items findet in der RD-Studie eher – wenn auch marginal – durch Workload mit 7.11 % und durch Gebrauchstauglichkeit mit 9.7 % statt. Der Anteil an diskriminanter Validität ist für die SAGAT-Items somit höher als der Anteil an konvergenter Validität. Umgekehrt verhält es sich in der 3DA Untersuchung. Hier ist der Anteil an konvergenter Validität höher als der der diskriminanten Validität. Die Items des SAGAT werden also durch die Verwendung der anderen Methoden, die zur Erfassung des Situationsbewusstseins genutzt wurden, besser erklärt (6.63 %) als durch die Methoden der Workload (5.84 %) oder der Gebrauchstauglichkeit (5.75 %). Insgesamt gesehen sind daher die Kriterien des SAGATs in der 3DA-Studie besser geeignet, um das Konstrukt Situationsbewusstsein zu erfassen als in der RD-Studie.

Tabelle 7.20 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs in Kombination mit Workload

Spearman's ρ in %	RD	3DA
$ \rho \approx 0$	65.61	61.36
$0.2 < \rho \leq 0.4$	27.27	32.79
$ \rho > 0.4$	7.11	5.84

Tabelle 7.21 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs in Kombination mit Gebrauchstauglichkeit

	RD	3DA
$ \rho \approx 0$	59.80	63.69
$0.2 < \rho \leq 0.4$	30.51	30.56
$ \rho > 0.4$	9.70	5.75

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

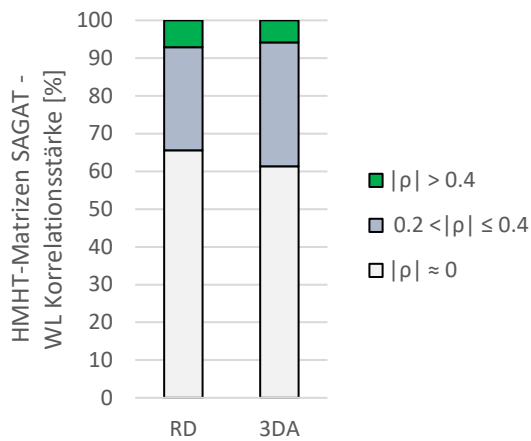


Abbildung 7.21 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Workload

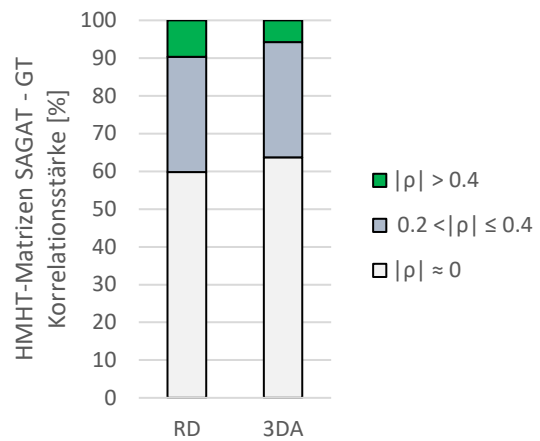


Abbildung 7.22 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Gebrauchstauglichkeit

7.2.1.3 Sortieraufgabe

Die Sortieraufgabe ist lediglich im Rahmen der Evaluierung der re-designten Navigationshilfe als funktionsrelevantes Maß eingeführt und genutzt worden. Eine Gegenüberstellung der Matrizen (vgl. Tabelle 7.22 und Abbildung 7.23) zeigt einen Anteil von 100 % bedeutsamen Korrelationen in der MMT-Matrix und einen eher geringen Anteil an konvergenter Validität von 8.60 % mit Blick auf die HMHT-Matrix.

Tabelle 7.22 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Sortieraufgabe

Spearman's ρ in %	MMMT	HMHT	HMHT -WL	HMHT -GT
$ \rho \approx 0$	0.00	65.59	63.77	42.96
$0.2 < \rho \leq 0.4$	0.00	25.81	27.54	37.78
$ \rho > 0.4$	100.00	8.60	8.70	19.26

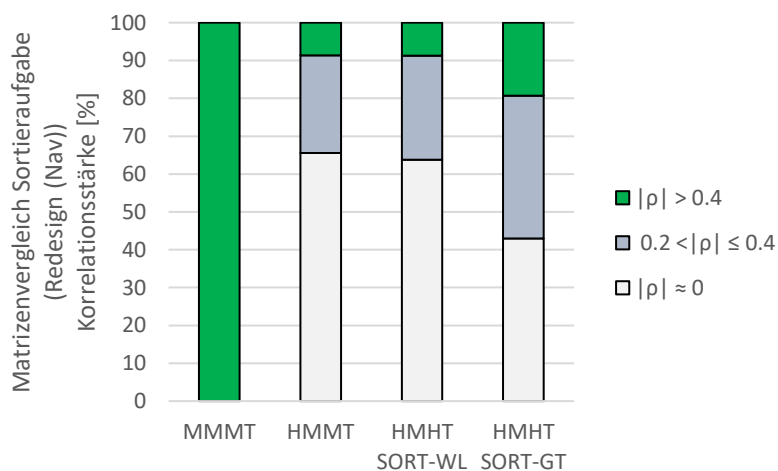


Abbildung 7.23 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Sortieraufgabe

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Ein Vergleich der HMMT-Matrix mit den HMHT-Matrizen verdeutlicht zudem, dass die Sortieraufgabe deutlich besser mit den Methoden der Gebrauchstauglichkeit korreliert (19.26 %) als mit den Methoden, die das Situationsbewusstsein erfassen sollen (8.6 %). Im Hinblick auf die Kombination der Sortieraufgabe mit Workload ist nahezu kein Unterschied zwischen der konvergenten Validität (8.6 %) und der diskriminanten Validität (8.7 %) erkennbar.

7.2.1.4 Gefahren-Detektierung

In der Evaluierung eines räumlich-auditiven Warnsignals ist ein funktionsrelevantes Vorgehen integriert worden, dass die Entdeckungsleistung der Probanden hinsichtlich einer Gefahrenquelle misst. Da dieses Vorgehen die perzeptuellen Leistungen der Probanden während der Situation beschreibt, wird dieses Vorgehen in der Matrix den Methoden zur Erfassung des Situationsbewusstseins zugeordnet und die Ergebnisse in Tabelle 7.23 und in Abbildung 7.24 gegenübergestellt. Im Hinblick auf die MMMT-Matrix zeigt sich, dass die eingeflossene Werte differenzieren, da der Anteil an bedeutsamen Korrelationen bei 50 % liegt.

Tabelle 7.23 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Gefahren-Detektierung

Spearman's ρ in %	MMMT	HMMT	HMHT -WL	HMHT -GT
$ \rho \approx 0$	16.67	45.83	37.50	43.75
$0.2 < \rho \leq 0.4$	33.33	46.88	39.77	35.42
$ \rho > 0.4$	50.00	7.29	22.73	20.83

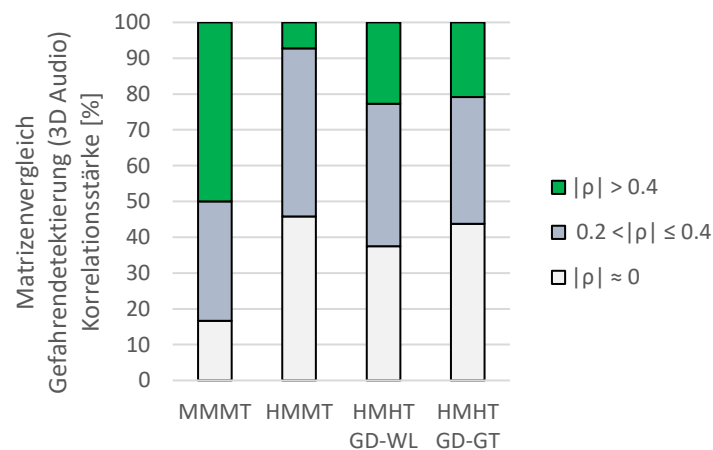


Abbildung 7.24 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Gefahren-Detektierung

Bei Überprüfung der Aufklärung der Gefahren-Detektierung innerhalb der Situationsbewusstseinsmethoden zeigt sich mit 7.29 % ein kleiner Anteil an bedeutsamen Korrelationen. Deutlicher ist die Verbindung dieser Methodik zum Konstrukt Gebrauchstauglichkeit mit 20.83 % und insbesondere zu den Methoden der Workload mit 22.73 %.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

7.2.1.5 Eye-Tracking

Zur Beschreibung der Aufmerksamkeit und deren Allokation wurden verschiedene Eye-Tracking-Maße wie beispielsweise Fixationen und deren jeweilige Dauer in die Matrix für den Bereich Situationsbewusstsein aufgenommen. Die Analyse der MMMT-Matrizen (vgl. Tabelle 7.24 und Abbildung 7.25) zeigen für diese Methodik einen stark schwankenden Anteil an bedeutsamen Korrelationen. Der geringste Anteil ist hier in der RD-Studie mit 26.67 %, der größte Anteil hingegen in der Machbarkeitsstudie mit 83.33 %. Die beiden anderen Studien sind dazwischen angesiedelt mit 46.67 % in der 3DA-Studie und 66.67 % in der Evaluierung der Sensoreinstellung.

Tabelle 7.24 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	33.33	13.33	20.00
$0.2 < \rho \leq 0.4$	16.67	40.00	40.00	13.33
$ \rho > 0.4$	83.33	26.67	46.67	66.67

Tabelle 7.25 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit

	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	55.95	59.09	34.44
$0.2 < \rho \leq 0.4$	0.00	27.38	35.61	32.22
$ \rho > 0.4$	100.00	16.67	5.30	33.33

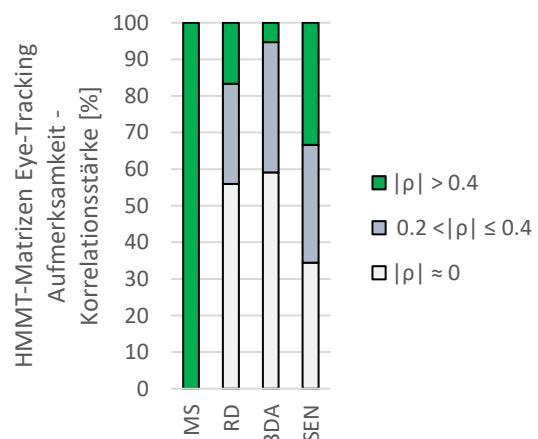
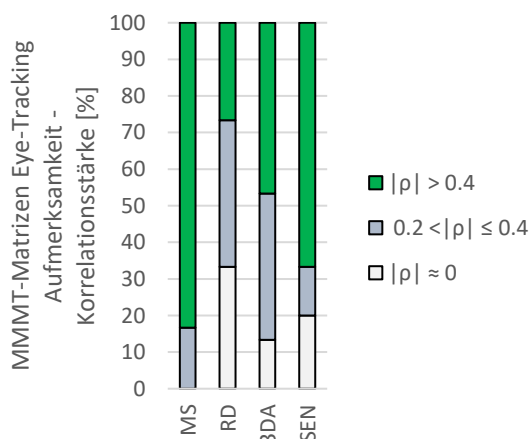


Abbildung 7.25 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit

Abbildung 7.26 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit

Bei Gegenüberstellung der HMMT-Matrizen (vgl. Tabelle 7.25 und Abbildung 7.26) und der HMHT-Matrizen (Tabelle 7.26 und Abbildung 7.27 sowie Tabelle 7.27 und Abbildung 7.28) sticht erneut die Machbarkeitsstudie mit einem Anteil von 100 % bedeutsamen Korrelationen hervor. Mit Bezug auf die Evaluation des Re-Designs und der Studie zur Untersuchung der Sensoreinstellungen stellt sich heraus, dass der Anteil bedeutsamer Korrelationen in den HMMT-Matrizen (RD = 16.67 %, SEN = 33.33 %) höher ist als der Anteil in den HMHT-Matrizen in Kombination mit Workload (RD = 15.22 %, SEN = 31.48 %). Wird die HMMT-Matrix der SEN-Studie (33.33 %) mit der SB-GT-HMHT-Matrix (44.87 %) verglichen, zeigt sich für die SEN-Studie, dass die Eye-Tracking-Maße die Gebrauchstauglichkeit besser erklären. Auch der Vergleich in der RD-Studie würde diesem Ergebnis entsprechen, jedoch ist der Unterschied zwischen konvergenter und diskriminanter Validität in diesem Fall vernachlässigbar.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Eine Ausnahme bildet die 3DA-Evaluation, da hier der Anteil an diskriminanter Validität immer deutlich höher als der der konvergenten Validität (5.30 %) ist. Demnach beschreiben die verwendeten Eye-Tracking-Maße Workload in der 3DA-Studie mit 24.24 %. Die diskriminante Validität im Hinblick auf die Kombination mit Gebrauchstauglichkeit ist hingegen mit 20.37 % etwas geringer.

Tabelle 7.26 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Workload

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	41.07	49.28	41.67	45.37
$0.2 < \rho \leq 0.4$	30.36	35.51	34.09	23.15
$ \rho > 0.4$	28.57	15.22	24.24	31.48

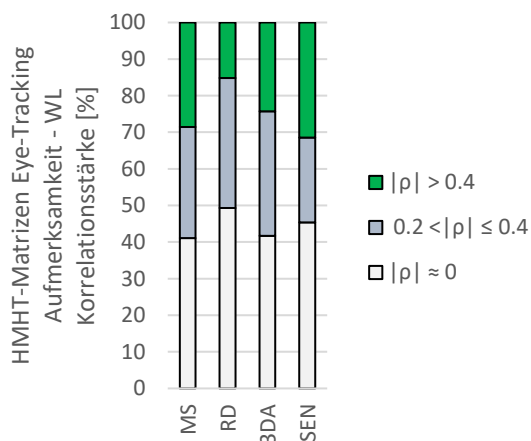


Abbildung 7.27 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Workload

Tabelle 7.27 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit

MS	RD	3DA	SEN
34.56	51.48	37.96	29.49
27.94	31.48	41.67	25.64
37.50	17.04	20.37	44.87

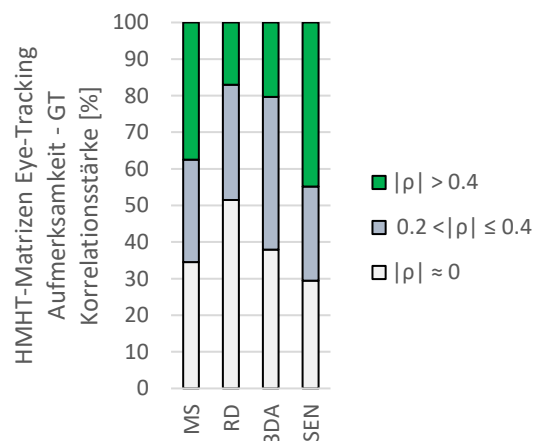


Abbildung 7.28 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit

Neben der globalen Aufmerksamkeit ist auch die gerichtete Aufmerksamkeit während der Bedienung der evaluierten technischen Funktionen erhoben worden. Mit Blick auf die MMMT-Matrizen (vgl. Tabelle 7.28 und Abbildung 7.29) zeigen sich hohen Anteile an bedeutsamen Korrelationen für die Machbarkeitsstudie, die Evaluierung einer Navigationshilfe sowie die Sensorstudie, da alle Matrizen deutlich über den 50 % liegen.

Tabelle 7.28 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit

Spearman's ρ in %	MS	RD	SEN
$ \rho \approx 0$	0.00	15.56	23.64
$0.2 < \rho \leq 0.4$	0.00	15.56	21.82
$ \rho > 0.4$	100.00	68.89	54.55

Tabelle 7.29 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit

MS	RD	SEN
0.00	51.67	24.55
0.00	35.42	30.91
100.00	12.92	44.55

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

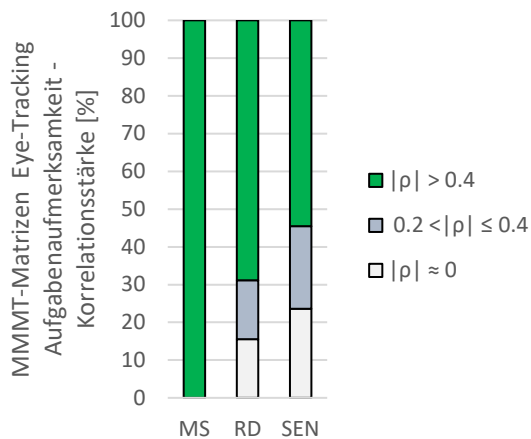


Abbildung 7.29 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit

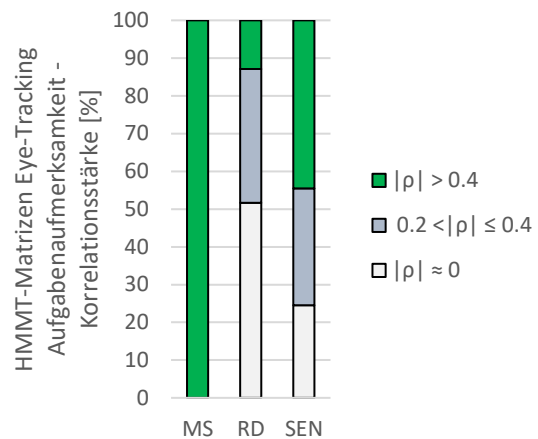


Abbildung 7.30 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit

Für einen Vergleich der HMMT-Matrizen (vgl. Tabelle 7.29 und Abbildung 7.30) mit den HMHT-Matrizen (vgl. Tabelle 7.30 und Abbildung 7.31 sowie Tabelle 7.31 und Abbildung 7.32) bezüglich der aufgabenbasierten Eye-Tracking-Messung bildet die Machbarkeitsstudie erneut eine Ausnahme, da hier zwei Eye-Tracking-Herangehensweisen in die HMMT-Matrix eingeflossen sind und zusammen den Wert von 100 % an bedeutsamen Korrelationen bedingen. Betrachtet man die anderen beiden Studien, sind die zu treffenden Aussagen eher uneindeutig. Der Anteil an konvergenter Validität ist in der RD-Studie mit 12,92 % deutlich geringer als der Anteil an bedeutsamen Korrelationen bei Kombination der aufgabenbasierten Eye-Tracking-Messung mit Workload (26,96 %) als auch in der Kombination mit Gebrauchstauglichkeit (26,22 %). Der Unterschied zwischen diesen beiden HMHT-Matrizen ist allerdings vernachlässigbar, da keines der beiden Konstrukte mit dieser Methodik eindeutig besser beschrieben wird. In der SEN-Evaluierung hingegen beschreibt die verwendete Eye-Tracking-Methode sowohl Situationsbewusstsein mit 44,55 % als auch Gebrauchstauglichkeit ähnlich gut (43,36 %). Die HMHT-Matrix weist allerdings einen etwas höheren Anteil an bedeutsamen Korrelationen in der SEN-Studie mit 48,48 % an bedeutsamen Korrelationen auf.

Tabelle 7.30 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Workload

Spearman's ρ in %	MS	RD	SEN
$ \rho \approx 0$	42,86	39,13	23,23
$0.2 < \rho \leq 0.4$	14,29	33,91	28,28
$ \rho > 0.4$	42,86	26,96	48,48

Tabelle 7.31 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit

	MS	RD	SEN
$ \rho \approx 0$	56,86	41,56	29,02
$0.2 < \rho \leq 0.4$	19,61	32,22	27,62
$ \rho > 0.4$	23,53	26,22	43,36

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

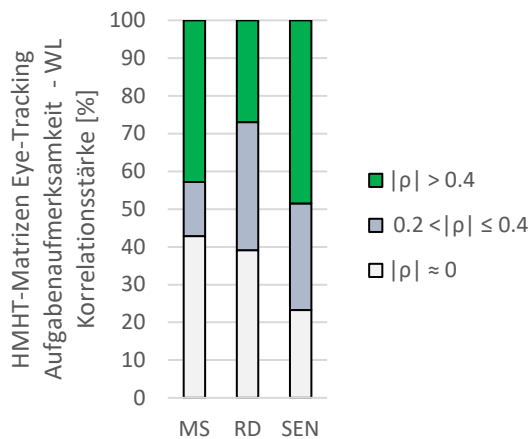


Abbildung 7.31 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Workload

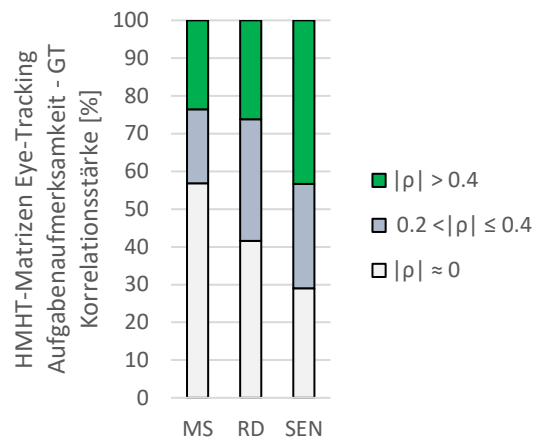


Abbildung 7.32 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit

7.2.2 Methoden zur Erfassung der Workload

Die Erhebung von Workload ist in den vier durchgeführten Evaluierungen anhand des subjektiven Fragebogens NASA TLX, der Anwendung des objektiven Messvorgehens TDRT sowie der Erfassung von unterschiedlichen Eye-Tracking-Maßen vorgenommen worden. Die Methoden sind iterativ und unter ökonomischen Aspekten in das Methoden-Inventar eingebunden worden.

7.2.2.1 NASA Task Load Index (NASA TLX)

Auch in der Anwendung des NASA TLX als vielfach verwendetem Standard-Verfahren zur Erhebung von Workload fällt insbesondere in der Machbarkeitsstudie der hohe Anteil an bedeutenden Korrelationen von 90.48 % in der MMT-Matrix auf (vgl. Tabelle 7.32 und Abbildung 7.33). Die 3DA-Studie und die SEN-Studie liegen hingegen bei jeweils bei 52.38 % und weisen einen deutlich niedrigeren Anteil an bedeutsamen Korrelationen in den MMT-Matrizen auf. Mit 61.90 % an bedeutsamen Korrelationen in der RD-Studie liegt hier der Anteil etwas über der 3DA- und der SEN-Studie.

Tabelle 7.32 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen des NASA TLX

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	4.76	9.52	9.52	38.10
$0.2 < \rho \leq 0.4$	4.76	28.57	38.10	9.52
$ \rho > 0.4$	90.48	61.90	52.38	52.38

Tabelle 7.33 Prozentuale Angabe der Korrelationsstärke in den HMHT-Matrizen des NASA TLX

	MS	RD	3DA	SEN
$ \rho > 0.4$	22.45	43.75	65.71	32.47
$0.2 < \rho \leq 0.4$	30.61	41.07	19.05	23.38
$ \rho \approx 0$	46.94	15.18	15.24	44.16

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

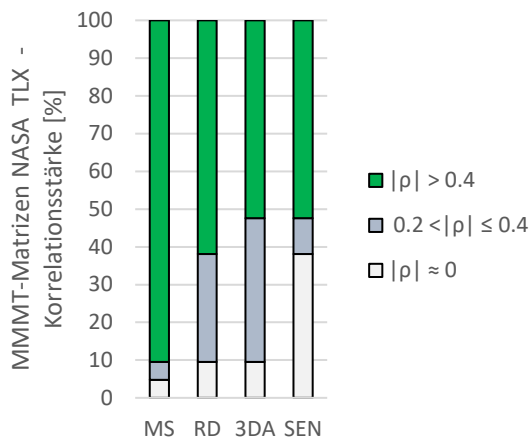


Abbildung 7.33 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des NASA TLX

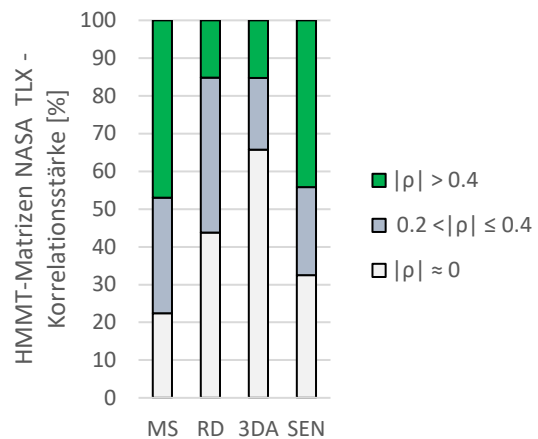


Abbildung 7.34 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des NASA TLX

Betrachtet man die Anteile der konvergenten Validität in den HMMT-Matrizen (vgl. Tabelle 7.33 und Abbildung 7.34) zeigt sich erneut, dass mit Verbreiterung des Methoden-Inventars und größerer Datenbasis der Anteil an bedeutsamen Korrelationen geringer ist (RD = 15.18 %, 3DA = 15.24 %). Da der Anteil an bedeutsamen Korrelationen in der Machbarkeitsstudie (46.94 %) und in der Evaluierung der Sensoreinstellungen deutlich höher liegt (44.16 %), heißt das, dass hier die Items deutlich enger miteinander und so mit dem Konstrukt zusammenhängen.

Stellt man nun die HMMT-Matrizen den HMHT-Matrizen (vgl. Tabelle 7.34 und Abbildung 7.35 sowie Tabelle 7.35 und Abbildung 7.36) gegenüber, zeigt sich noch mehr Ambivalenz. In der Machbarkeitsstudie erklärt sich der NASA TLX durch die Methoden der Workload (46.49 %) deutlich besser als durch die Methoden zur Erfassung des Situationsbewusstseins (26.53 %), jedoch schlechter, als wenn der NASA TLX mit den Methoden der Gebrauchstauglichkeit (70.17 %) korreliert wird. In der SEN-Studie hingegen ist der Anteil an konvergenter Validität (44.16 %) höher als bei Korrelation des NASA TLX mit Situationsbewusstsein (39.46 %) oder mit Gebrauchstauglichkeit (32.97 %). Hier zeigt sich auch im Gegensatz zu der Machbarkeitsstudie eine leicht engere Verwandtschaft mit dem Situationsbewusstsein, wohingegen in der Machbarkeitsstudie eine nähere Verbindung zur Gebrauchstauglichkeit besteht. Auch in der 3DA-Studie wird deutlich, dass der NASA TLX innerhalb des Konstruktes (15.24 %) häufiger höher korreliert als außerhalb des Konstruktes. Im Vergleich der HMHT-Matrizen scheint der NASA TLX marginal enger mit dem Situationsbewusstsein (10.20 %) verbunden zu sein als mit dem Konstrukt Gebrauchstauglichkeit (8.33 %). In der Untersuchung des Re-Designs hingegen ist wiederum der Anteil an bedeutsamen Korrelationen in den HMHT-Matrizen höher als in der HMMT-Matrix (15.18 %). In dieser Studie zeigt sich eine engere Beziehung des NASA TLX zum Konstrukt Situationsbewusstsein (24.37 %) als zu den anderen Workload-Methoden oder dem Konstrukt Gebrauchstauglichkeit (21.90 %).

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.34 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	36.73	45.80	58.16	31.97
$0.2 < \rho \leq 0.4$	36.73	29.83	31.63	28.57
$ \rho > 0.4$	26.53	24.37	10.20	39.46

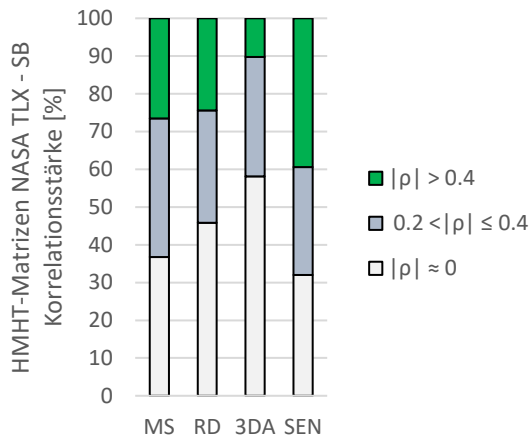


Abbildung 7.35 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Situationsbewusstsein

Tabelle 7.35 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Gebrauchstauglichkeit

	MS	RD	3DA	SEN
$ \rho \approx 0$	10.50	41.27	57.54	39.56
$0.2 < \rho \leq 0.4$	19.33	36.83	34.13	27.47
$ \rho > 0.4$	70.17	21.90	8.33	32.97

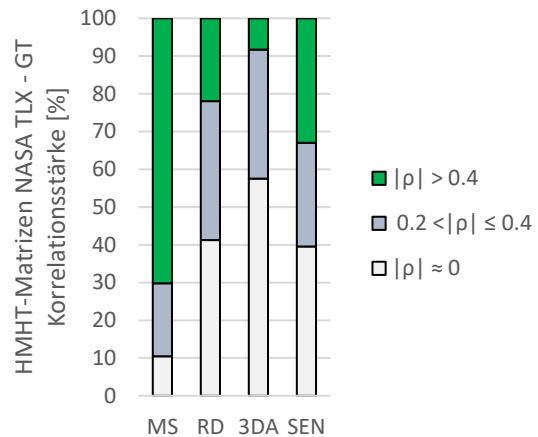


Abbildung 7.36 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Gebrauchstauglichkeit

7.2.2.2 Tactile Detection Response Task (TDRT)

Als objektives Messvorgehen wurde die TDRT zur Evaluierung des Re-Designs und in der Evaluierung des räumlich-auditiven Warnsignals genutzt. Bei Betrachtung der MMT-Matrizen (vgl. Tabelle 7.36 und Abbildung 7.37) sind zwei sehr unterschiedliche Ergebnisse ersichtlich. In der RD-Studie ist der Anteil an bedeutsamen Korrelationen mit 60 % relativ hoch, während der Anteil an bedeutsamen Korrelationen in der 3DA-Studie mit 16.67 % sogar sehr niedrig ist.

Tabelle 7.36 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der TDRT

Spearman's ρ in %	RD	3DA
$ \rho \approx 0$	0.00	66.67
$0.2 < \rho \leq 0.4$	40.00	16.67
$ \rho > 0.4$	60.00	16.67

Tabelle 7.37 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der TDRT

	RD	3DA
$ \rho \approx 0$	70.00	52.78
$0.2 < \rho \leq 0.4$	25.56	37.50
$ \rho > 0.4$	4.44	9.72

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

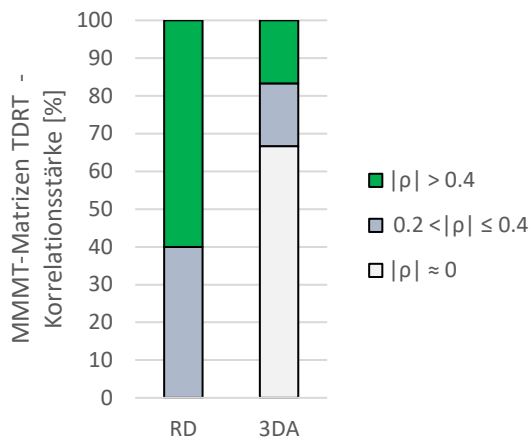


Abbildung 7.37 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der TDRT

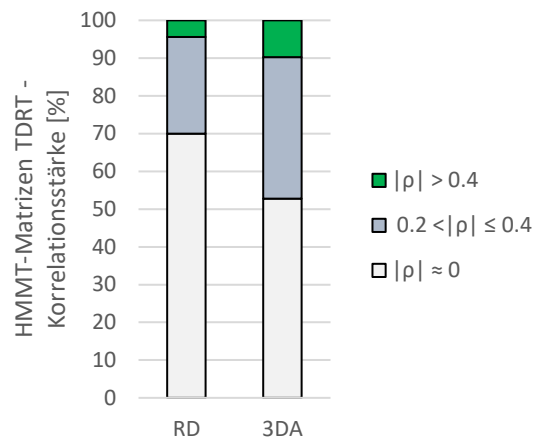


Abbildung 7.38 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der TDRT

Im Hinblick auf die konvergente Validität ergibt sich in beiden Studien eine geringe Verbundenheit der TDRT mit dem Konstrukt Workload und den hierin angewendeten Methoden (vgl. Tabelle 7.37 und Abbildung 7.38). So ist der Anteil an bedeutsamen Korrelationen in der 3DA-Studie mit 9,72 % niedrig; der Anteil in der RD-Studie ist mit 4,44 % jedoch äußerst gering. Verglichen hierzu sind die Anteile an bedeutsamen Korrelationen in den HMHT-Matrizen (vgl. Tabelle 7.38 und Abbildung 7.39 sowie Tabelle 7.39 und Abbildung 7.40) deutlich höher. In beiden Fällen ist die TDRT mit den Methoden zur Erfassung des Situationsbewusstseins (RD = 8,82 %; 3DA = 15,18 %) weniger stark zusammenhängend als mit den Methoden der Gebrauchstauglichkeit (RD = 14,22 %, 3DA = 18,06 %). Der Vergleich der Anteile an konvergenter Validität mit den Matrizen der diskriminanten Validität ergibt somit, dass die TDRT in beiden Fällen besser dazu geeignet wäre, Gebrauchstauglichkeit zu beschreiben als Situationsbewusstsein oder das eigene Konstrukt Workload.

Tabelle 7.38 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Situationsbewusstsein

Spearman's ρ in %	RD	3DA
$ \rho \approx 0$	64.12	48.21
$0.2 < \rho \leq 0.4$	27.06	36.61
$ \rho > 0.4$	8.82	15.18

Tabelle 7.39 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Gebrauchstauglichkeit

	RD	3DA
$ \rho \approx 0$	50.67	47.92
$0.2 < \rho \leq 0.4$	35.11	34.03
$ \rho > 0.4$	14.22	18.06

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

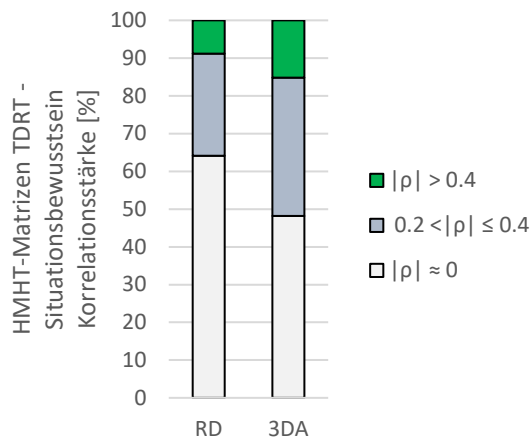


Abbildung 7.39 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Situationsbewusstsein

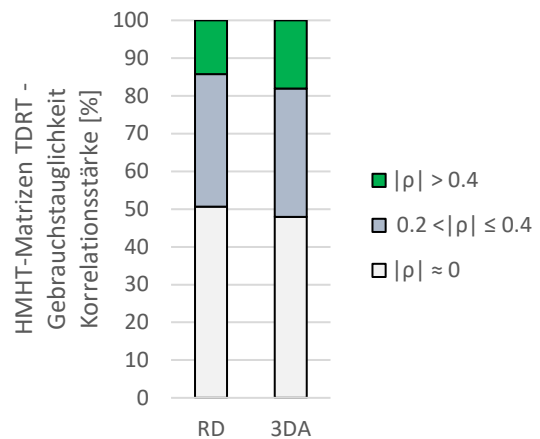


Abbildung 7.40 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Gebrauchstauglichkeit

7.2.2.3 Eye-Tracking

Ein Methodenansatz zur Bestimmung von Workload mittels Eye-Tracking ist die Erhebung der Pupillen-Dilatation. Da die erhobenen Werte der MMT-Matrizen (vgl. Tabelle 7.40 und Abbildung 7.41) in allen vier Evaluierungen jeweils zu 100 % bedeutsam korrelieren, wird deutlich, dass die innerhalb dieser Methode verwendeten Items sehr eng miteinander verbunden sind.

Tabelle 7.40 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	0.00	0,00	0.00
$0.2 < \rho \leq 0.4$	0.00	0.00	0,00	0.00
$ \rho > 0.4$	100.00	100.00	100.00	100.00

Tabelle 7.41 Prozentuale Angabe der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung

MS	RD	3DA	SEN
27.27	62.75	69.79	26.39
27.27	34.31	20.83	18.06
45.45	2.94	9.37	55.56

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

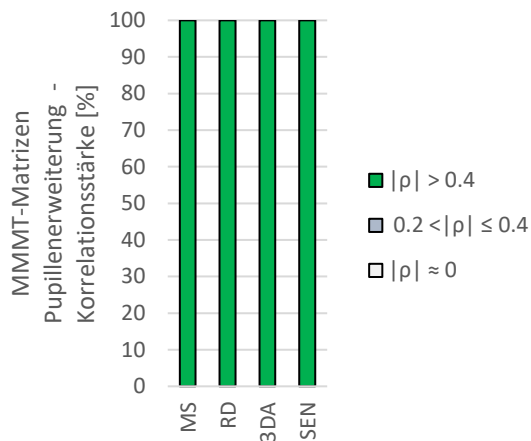


Abbildung 7.41 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung

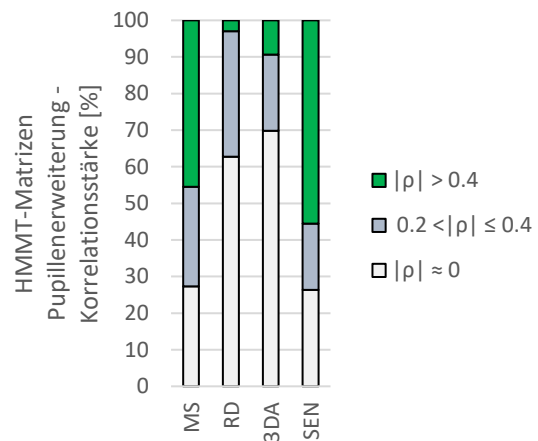


Abbildung 7.42 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung

Eine Gegenüberstellung der HMHT-Matrizen (vgl. Tabelle 7.41 und Abbildung 7.42) mit den MMT-Matrizen (vgl. Tabelle 7.42 und Abbildung 7.43 sowie Tabelle 7.43 und Abbildung 7.44) zeigt deutlich, dass die diskriminante Validität höher als die konvergente Validität ist. Lediglich in der Evaluierung der Sensorbedienung ist der Anteil an konvergenter Validität (55.56 %) höher als die Kombination der Pupillenerweiterung mit den Methoden zur Erfassung des Situationsbewusstseins (49.21 %) oder den Methoden der Gebrauchstauglichkeit (30.13 %). In der Machbarkeitsstudie wird das Situationsbewusstsein (71.43 %) mittels der Pupillenerweiterung besser erklärt als durch die Methoden, die zur Erhebung der Workload (45.45 %) genutzt wurden. Im Hinblick auf die Kombination der Pupillenerweiterung mit den Methoden der Gebrauchstauglichkeit (26.47 %) hängt die Pupillenerweiterung wiederum enger mit dem Konstrukt Workload zusammen. Den geringsten Anteil an bedeutsamen Korrelationen bezüglich der konvergenten Validität ist mit 2.94 % in der RD-Studie zu verzeichnen. Hier wäre die Methodik geeigneter gewesen, um die Methoden der Gebrauchstauglichkeit (13.33 %) zu erklären. Auch die Kombination mit den Methoden zur Erfassung des Situationsbewusstseins führt in dieser Studie mit 10.78 % zu einem deutlich höheren Wert in der HMHT-Matrix als im Vergleich hierzu in der MMT-Matrix. In der 3DA-Studie hingegen ist bezüglich der Pupillenerweiterung ein marginal höherer Wert im Hinblick auf die HMHT-Matrix in Kombination mit Situationsbewusstsein zu erkennen (14.29 %). Auch die HMHT-Matrix, welche die Zusammenhänge mit den Methoden der Gebrauchstauglichkeit beschreibt, verdeutlicht hier einen höheren Wert an bedeutsamen Korrelationen (13.89 %) als innerhalb des Konstruktes Workload (9.37 %).

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.42 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	28.57	53.92	47.02	23.02
$0.2 < \rho \leq 0.4$	0.00	35.29	38.69	27.78
$ \rho > 0.4$	71.43	10.78	14.29	49.21

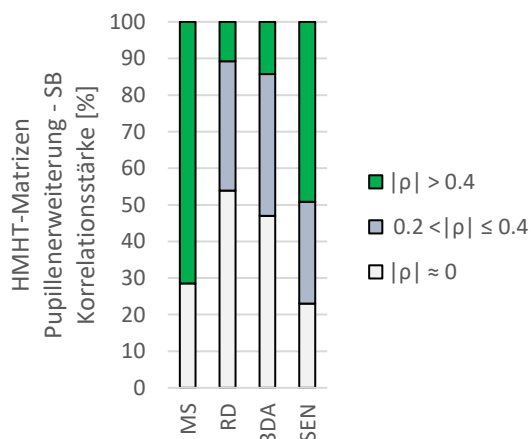


Abbildung 7.43 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Situationsbewusstsein

Ein weiterer Parameter, der mittels Eye-Tracking ermittelt werden kann und der zur Bestimmung von Müdigkeit geeignet ist, ist das Maß PERCLOS. Erneut wird eine deutliche Abhängigkeit der Messitems innerhalb der Methode für alle Studien offensichtlich. Jede der vier Studien in den MMT-Matrizen (vgl. Tabelle 7.44 und Abbildung 7.45) ist mit 100 % an bedeutsamen Korrelationen vertreten.

Tabelle 7.44 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung PERCLOS

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	0.00	0.00	0.00
$0.2 < \rho \leq 0.4$	0.00	0.00	0.00	0.00
$ \rho > 0.4$	100.00	100.00	100.00	100.00

Tabelle 7.43 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Gebrauchstauglichkeit

MS	RD	3DA	SEN
50.00	56.67	53.70	33.33
23.53	30.00	32.41	36.54
26.47	13.33	13.89	30.13

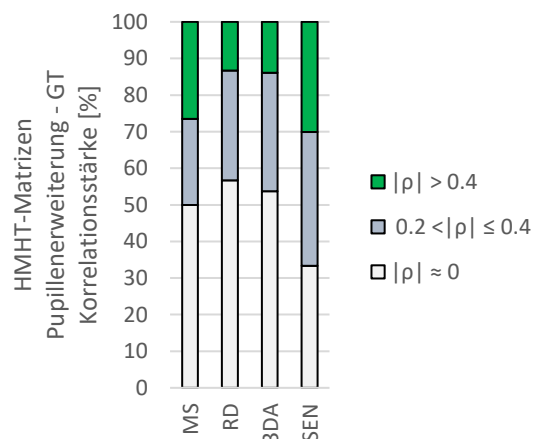


Abbildung 7.44 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Gebrauchstauglichkeit

Tabelle 7.45 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung PERCLOS

MS	RD	3DA	SEN
25.00	58.33	56.14	60.00
54.17	28.33	40.35	6.67
20.83	13.33	3.51	33.33

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

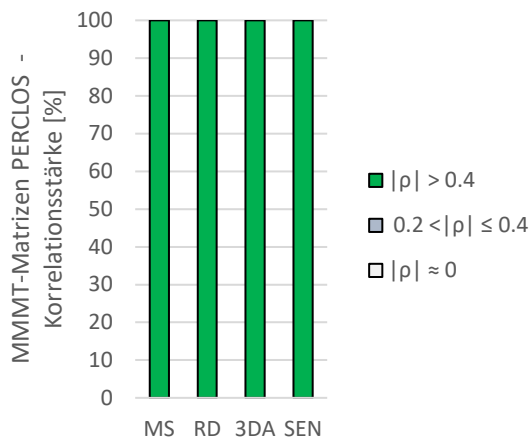


Abbildung 7.45 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung PERCLOS

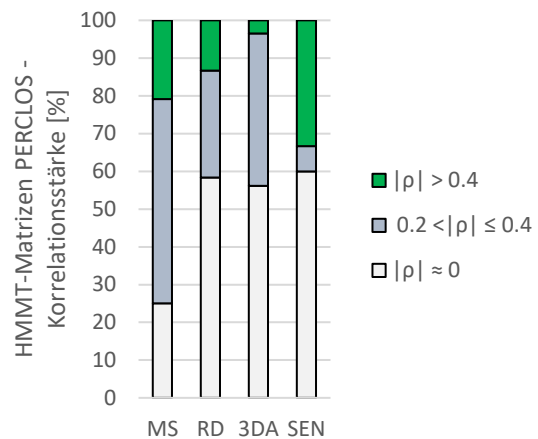


Abbildung 7.46 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung PERCLOS

Die Ergebnisse der Gegenüberstellung der HMMT-Matrizen (vgl. Tabelle 7.45 und Abbildung 7.46) gegenüber den HMHT-Matrizen (vgl. Tabelle 7.46 und Abbildung 7.47 sowie Tabelle 7.47 und Abbildung 7.48) deutet darauf hin, dass eine Abgrenzung der Konstrukte unter Verwendung des PERCLOS nicht möglich ist. Lediglich in der Machbarkeitsstudie sind die Anteile in der HMMT-Matrix mit 20.83 % höher als die HMHT-Matrix in Kombination mit Situationsbewusstsein, die keinerlei bedeutsame Korrelationen aufweist. Dennoch wäre auch in dieser Studie der PERCLOS eher dazu geeignet gewesen, die Methoden der Gebrauchstauglichkeit (30.88 %) zu erklären als Workload (20.38 %). In der RD-Studie und der 3DA-Evaluierung sind die Anteile an diskriminanter Validität in der Kombination mit den Methoden zur Erfassung des Situationsbewusstseins höher (RD = 17.65 %, 3DA = 15.48 %) als in Verbindung mit Gebrauchstauglichkeit (RD = 14.81 %, 3DA = 12.04 %). In beiden Fällen jedoch ist der Anteil an bedeutsamen Korrelationen in den HMMT-Matrizen niedriger (RD = 13.33 %, 3DA = 3.51 %). Das gilt auch für die Evaluierung der Sensorbedienung mit einem Wert von 33.33 % an bedeutsamen Korrelationen in der HMMT-Matrix. Im Rahmen dieser Evaluierung werden allerdings auch die Methoden der Gebrauchstauglichkeit (50 %) marginal besser erklärt als die Methoden zur Erfassung des Situationsbewusstseins mit 47.62 %.

Tabelle 7.46 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	78.57	52.94	52.38	33.33
$0.2 < \rho \leq 0.4$	21.43	29.41	32.14	19.05
$ \rho > 0.4$	0.00	17.65	15.48	47.62

Tabelle 7.47 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Gebrauchstauglichkeit

	MS	RD	3DA	SEN
$ \rho > 0.4$	22.06	37.04	57.41	38.46
$0.2 < \rho \leq 0.4$	47.06	48.15	30.56	11.54
$ \rho \approx 0$	30.88	14.81	12.04	50.00

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

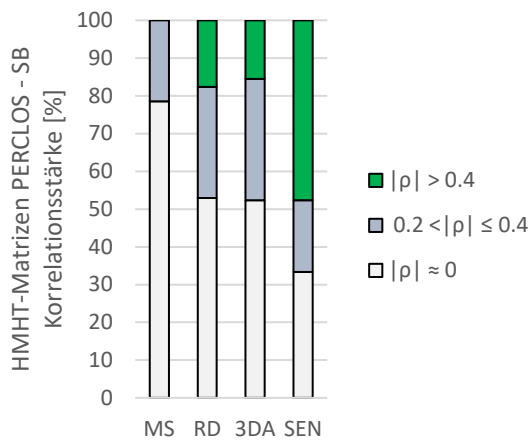


Abbildung 7.47 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Situationsbewusstsein

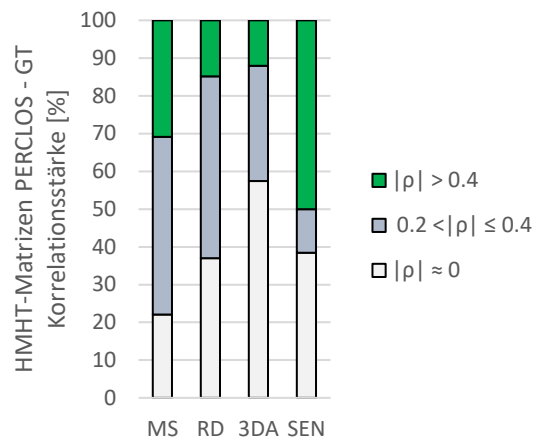


Abbildung 7.48 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Gebrauchstauglichkeit

Als weitere Operationalisierung der Workload wurde das Konzept der Aufmerksamkeitsfixierung als Methodik in das Methoden-Inventar eingebunden. Hierzu wurden überlange Fixationen und Blicke zu einer Methode zusammengefasst und mit Hilfe des multidimensionalen Methoden-Inventars validiert. Bei Betrachtung der MMT-Matrizen (vgl. Tabelle 7.48 und Abbildung 7.49) aller vier Studien ist daher auffällig, dass sich die verwendeten Messwerte innerhalb der Methode uneinheitlich verhalten. Sowohl die Machbarkeitsstudie als auch die Untersuchung des räumlich-auditiven Warnsignals beinhalten keinerlei bedeutsame oder geringe Korrelationen. Die Studie zur Untersuchung der Sensorbedienung hingegen weist lediglich 100 % marginale Korrelationen auf, während in der Untersuchung des Re-Designs die Mess-Items zu 100 % bedeutsam korrelieren.

Tabelle 7.48 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	100.00	0.00	100.00	0.00
$0.2 < \rho \leq 0.4$	0.00	0.00	0.00	100.00
$ \rho > 0.4$	0.00	100.00	0.00	0.00

Tabelle 7.49 Prozentuale Angabe der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung

	MS	RD	3DA	SEN
$ \rho \approx 0$	16.67	64.29	80.00	53.13
$0.2 < \rho \leq 0.4$	29.17	30.95	15.00	25.00
$ \rho > 0.4$	54.17	4.76	5.00	21.88

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

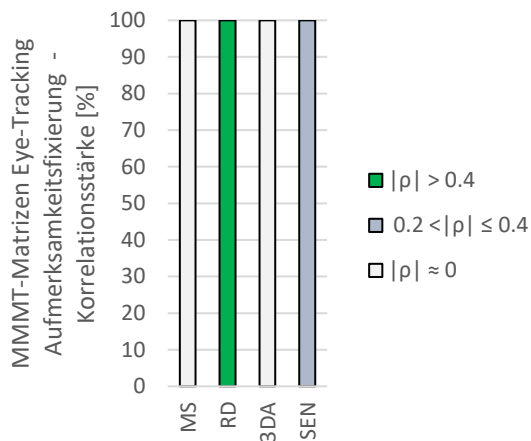


Abbildung 7.49 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung

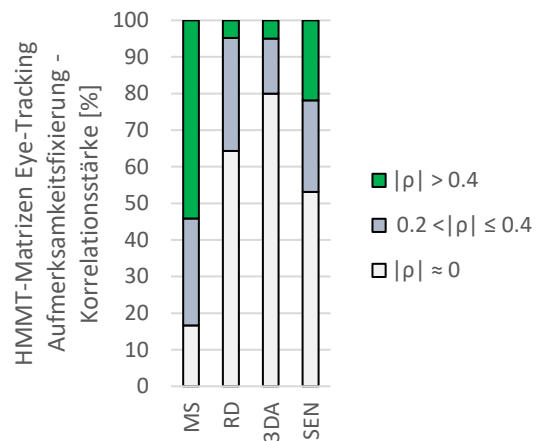


Abbildung 7.50 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung

Werden die HMMT-Matrizen (vgl. Tabelle 7.49 und Abbildung 7.50) analysiert, weisen die Machbarkeitsstudie und die SEN-Studie einen höheren Anteil an bedeutsamen Korrelationen auf (MS = 54.17 %, SEN = 21.88 %). In den anderen beiden Studien mit breiterer Datenbasis wird hingegen ein deutlich geringerer Anteil an bedeutsamen Korrelationen offensichtlich (RD = 4.76 %, 3DA = 5 %). Vergleicht man die HMMT-Matrizen mit den HMHT-Matrizen (vgl. Tabelle 7.50 und Abbildung 7.51 sowie Tabelle 7.51 und Abbildung 7.52), stellt sich lediglich in der Machbarkeitsstudie heraus, dass die Aufmerksamkeitsfixierung stärker mit dem Konstrukt Workload (54.17 %) korreliert als mit den Methoden zur Erfassung des Situationsbewusstseins (42.86 %). Allerdings werden in der MS-Studie mittels dieser Methodik die Methoden der Gebrauchstauglichkeit ähnlich gut erklärt (54.41 %). In den verbleibenden Evaluierungen sind die Anteile an konvergenter Validität niedriger als in den HMHT-Matrizen. In den Untersuchungen des Re-Designs und der 3DA-Evaluierung werden die Methoden der Gebrauchstauglichkeit (RD = 42.22 %; 3DA = 11.11 %) mittels der Aufmerksamkeitsfixierung besser beschrieben als die Methoden zur Erfassung des Situationsbewusstseins (RD = 17.65 %, 3DA = 8.93 %). In der SEN-Studie hingegen werden die Methoden zur Erfassung des Situationsbewusstseins (50 %) marginal besser erklärt als die Methoden der Gebrauchstauglichkeit (48.08 %).

Tabelle 7.50 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	42.86	48.53	71.43	30.95
$0.2 < \rho \leq 0.4$	14.29	33.82	19.64	19.05
$ \rho > 0.4$	42.86	17.65	8,93	50.00

Tabelle 7.51 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Gebrauchstauglichkeit

	MS	RD	3DA	SEN
	23.53	32.22	56.94	26.92
	22,06	25.56	31.94	25.00
	54.41	42.22	11.11	48.08

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

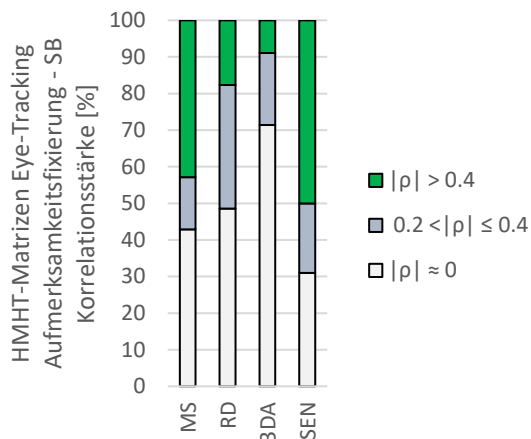


Abbildung 7.51 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Situationsbewusstsein

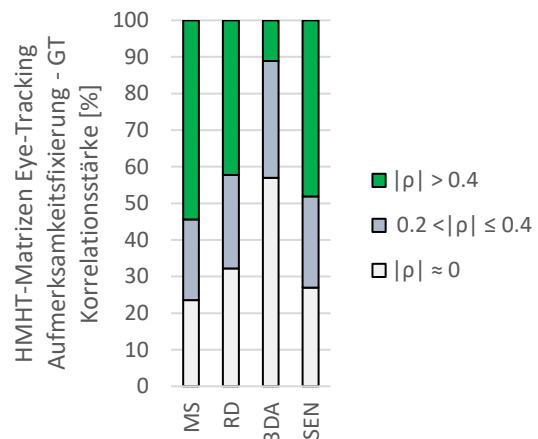


Abbildung 7.52 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Gebrauchstauglichkeit

7.2.3 Methoden zur Erfassung der Gebrauchstauglichkeit

Das Methoden-Inventar zur Erhebung der Gebrauchstauglichkeit beinhaltet neben dem subjektiven Fragebogen PSSUQ zur Erhebung der Nutzerzufriedenheit mehrere objektive Messvorgehen wie Bedienzeiten sowie Eye-Tracking-Messung als auch verschiedene Aspekte, die aus den Simulator-Aufzeichnungen abgeleitet wurden.

7.2.3.1 Nutzerleistung

Als klassisches Instrument zur Erfassung der Gebrauchstauglichkeit wurden verschiedene Nutzerleistungen erfasst. Hierunter fallen Zeitleistungen wie die Geschwindigkeit, mit der eine bestimmte Aufgabe des Nutzers erfüllt wurde, aber auch Leistungsmaße, welche die Genauigkeit der Aufgabenerfüllung erheben. Bei Analyse der Zeitleistungs-MMMT-Matrizen (vgl. Tabelle 7.52 und Abbildung 7.53) ist mit Ausnahme der Evaluierung der Sensoreinstellung ein großer Anteil bedeutsamer Korrelationen zu konstatieren (MS = 100 %, 3DA = 41.67 %). Da in der RD-Studie lediglich die Bedienzeit während der Nutzung der Navigationshilfe erfasst wurde, wurde der Wert vernachlässigt, da Eigenkorrelationen ausgeschlossen wurden. In der SEN-Evaluierung konnten lediglich marginal relevante Korrelationen nachgewiesen werden (SEN = 100 %).

Tabelle 7.52 Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Zeitleistung

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00		19.44	0.00
$0.2 < \rho \leq 0.4$	0.00		38.89	100.00
$ \rho > 0.4$	100.00		41.67	0.00

Tabelle 7.53 Prozentuale Angabe der Korrelationsstärken in den HMHT-Matrizen der Zeitleistung

	MS	RD	3DA	SEN
$ \rho > 0.4$	14.06	34.09	45.27	31.25
$0.2 < \rho \leq 0.4$	17.19	45.45	39.92	45.83
$ \rho \approx 0$	68.75	20.45	14.81	22.92

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

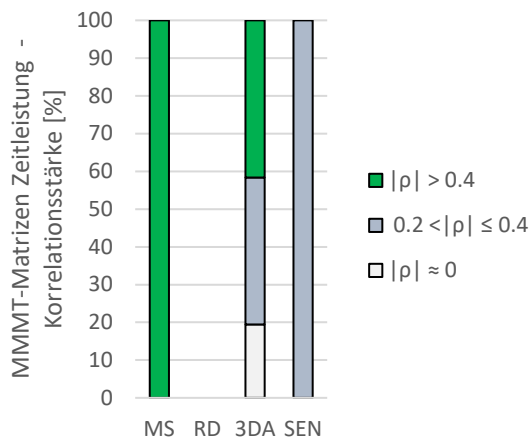


Abbildung 7.53 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Zeitleistung

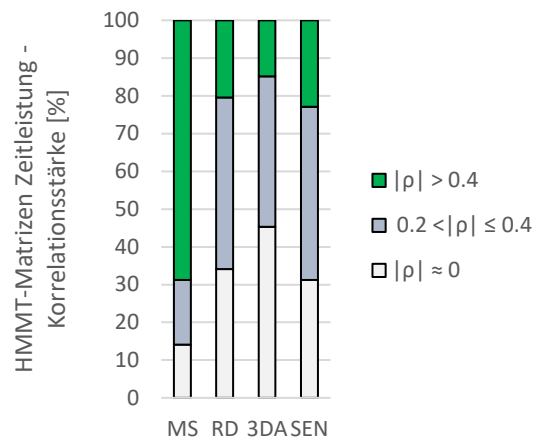


Abbildung 7.54 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Zeitleistung

Betrachtet man die HMMT-Matrizen (vgl. Tabelle 7.53 und Abbildung 7.54), zeigen sich für die Zeitleistung relativ hohe Korrelationen – auch in den Studien mit breiterer Datenbasis. Insbesondere in der Machbarkeitsstudie ist der Anteil an konvergenter Validität mit 68.75 % besonders hoch. Immer noch hoch, jedoch im Vergleich zur Machbarkeitsstudie niedrig, sind die Anteile an bedeutsamen Korrelationen in der SEN-Studie mit 22.92 %, der RD-Studie mit 20.45 % und schließlich der 3DA Evaluierung mit 14.81 %. Im Vergleich mit den HMHT-Matrizen (vgl. Tabelle 7.54 und Abbildung 7.55 sowie Tabelle 7.55 und Abbildung 7.56) zeigt sich lediglich für die Untersuchung des räumlich-auditiven Warnsignals eine klare Zuordnung der Zeitleistung zu dem Konstrukt Gebrauchstauglichkeit, da hier die Anteile an bedeutsamen Korrelationen in der HMHT-Matrix in Kombination mit Situationsbewusstsein (13.89 %) und die HMHT-Matrix in Kombination mit Workload (9.09 %) etwas niedriger sind. Uneindeutig ist hingegen das Ergebnis der Machbarkeitsstudie. Die Workload-Methoden (71.43 %) werden anhand der Zeitleistung marginal besser beschrieben als die Methoden der Gebrauchstauglichkeit (68.75 %). Jedoch besteht eine engere Verbindung zu dem Konstrukt Gebrauchstauglichkeit als zu dem Konstrukt Situationsbewusstsein und dessen Methoden (21.43 %). In der SEN-Evaluierung sind die Anteile an diskriminanter Validität hingegen sowohl in der Kombination mit den Methoden zur Erfassung des Situationsbewusstseins (40.48 %) als auch in Verbindung mit den Workload-Methoden (30.56 %) höher. Umgekehrt verhält es sich stattdessen bei Analyse der Ergebnisse in der RD-Evaluierung. In diesem Fall werden wiederum die Methoden der Workload (39.13 %) mittels Zeitleistung besser beschrieben als die Methoden zur Erfassung des Situationsbewusstseins (29.41 %). Beide HMHT-Matrizen der RD-Studie weisen allerdings höhere Anteile an diskriminanter Validität als Anteile an konvergenter Validität (20.45 %) auf.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.54 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Zeitleistung in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	57.14	44.12	45.63	21,43
$0.2 < \rho \leq 0.4$	21.43	26.47	40.48	38,10
$ \rho > 0.4$	21.43	29.41	13.89	40,48

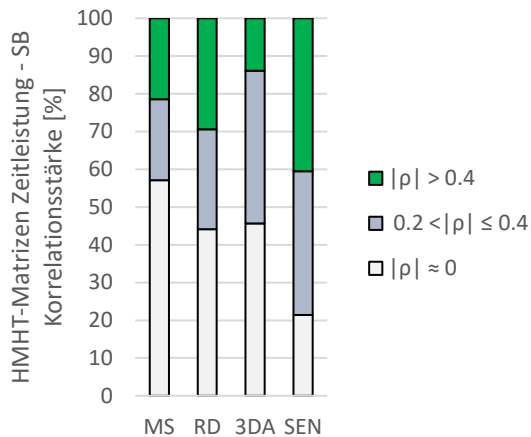


Abbildung 7.55 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Reaktionszeiten in Kombination mit Situationsbewusstsein

Tabelle 7.55 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Zeitleistung in Kombination mit Workload

	MS	RD	3DA	SEN
	28.57	34.78	53.54	36.11
	0.00	26.09	37.37	33.33
	71.43	39.13	9.09	30.56

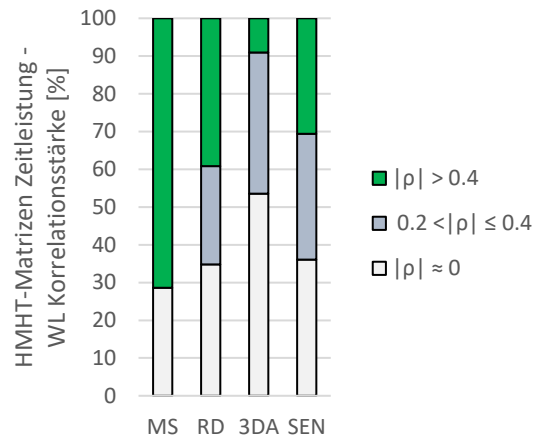


Abbildung 7.56 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Reaktionszeiten in Kombination mit Workload

Neben den Bedienzeiten wurden auch Simulator-Messungen – vor allem in der Untersuchung des räumlich-auditiven Warnsignals – abgeleitet und zur Bestimmung der Nutzerleistung genutzt. Diese Messungen wurden in erster Linie zur Bestimmung des Genauigkeitsgrades während der Aufgabenerfüllung verwendet.

Mitunter wurde in der 3DA-Studie erfasst, wie gut sich der Proband während der Durchführung seiner Aufgabe an die Instruktion halten konnte, eine bestimmte Geschwindigkeitsmaßgabe einzuhalten. Eine Gegenüberstellung der Matrizen (vgl. Tabelle 7.56 und Abbildung 7.57) zeigt zum einen 100 % bedeutsame Korrelationen in der MMT-Matrix. Wenn auch der Anteil an bedeutsamen Korrelationen in den verbleibenden Matrizen eher gering ist, so ist doch zum anderen zu erkennen, dass in der 3DA-Studie die Geschwindigkeitsleistung am besten dazu geeignet ist, um Gebrauchstauglichkeit zu erklären (7.35 %). Im Vergleich hierzu werden sowohl die Methoden zur Erfassung des Situationsbewusstseins (5.36 %) als auch die Methoden der Workload (4.55 %) weniger gut beschrieben.

Tabelle 7.56 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Geschwindigkeitsleistung in der 3DA-Studie

Spearman's ρ in %	MMMT	HMMT	HMHT-SB	HMHT-WL
$ \rho \approx 0$	0.00	61.76	62.50	56.82
$0.2 < \rho \leq 0.4$	0.00	30.88	32.14	38.64
$ \rho > 0.4$	100.00	7.35	5.36	4.55

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

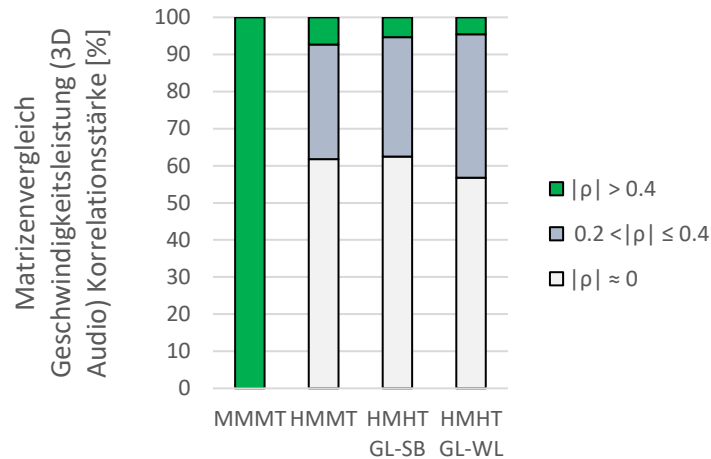


Abbildung 7.57 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Geschwindigkeitsleistung in der 3DA-Studie

Eine weitere Genauigkeitsmessung der 3DA-Studie beinhaltete die Genauigkeit der Ausrichtung des Flugzeuges im Raum während eines bestimmten Flugmanövers. Die MMMT-Matrix zeigt einen Anteil an bedeutsamen Korrelationen von 46.67 %. Deutlich wird weiter, dass anhand dieser Methodik das Konstrukt Gebrauchstauglichkeit in der 3DA-Studie gut von den Methoden zur Erfassung des Situationsbewusstseins (9.52 %) und der Workload (12.88 %) abgegrenzt werden konnte. Der Anteil an bedeutsamen Korrelationen für konvergente Validität in der HMMT-Matrix ist folglich mit 18.89 % deutlich höher. Die Methoden der Workload hängen zudem etwas stärker mit der Messung der Winkelgenauigkeit zusammen als die Methoden des Situationsbewusstseins. Ein Überblick der Ergebnisse der 3DA-Studie mit dem zugehörigen Matrizenvergleich ist in Tabelle 7.57 und Abbildung 7.58 zu finden.

Tabelle 7.57 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Winkelgenauigkeit (Notch) in der 3DA-Studie

Spearman's ρ in %	MMMT	HMMT	HMHT-SB	HMHT-WL
$ \rho \approx 0$	46.67	52.22	55.95	51.52
$0.2 < \rho \leq 0.4$	6.67	28.89	34.52	35.61
$ \rho > 0.4$	46.67	18.89	9.52	12.88

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

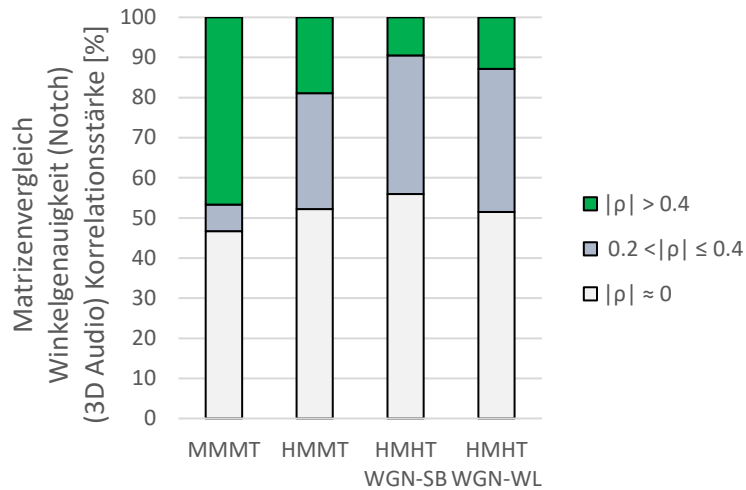


Abbildung 7.58 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Winkelgenauigkeit (Notch) in der 3DA-Studie

Ein zusätzliches Maß der Gebrauchstauglichkeit waren in der 3DA-Studie die von den Probanden durchgeführten Steuereingaben und deren Frequenz. Die MMMT-Matrix zeigt hier mit 100 % bedeutsamen Korrelationen eine hohe Abhängigkeit der Mess-Items untereinander auf. Die Gegenüberstellung von konvergenter und diskriminanter Validität (vgl. Tabelle 7.58 und Abbildung 7.59) ergibt, dass die Erhebung der Steuereingaben in der 3DA-Evaluierung geeignet war, um Gebrauchstauglichkeit gut zu beschreiben, da der Anteil an bedeutsamen Korrelationen 17.65 % ist. Die Methoden der Workload werden mit 9.09 % weniger gut erklärt als die Methoden zur Erfassung des Situationsbewusstseins mit 3.57 %.

Tabelle 7.58 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Steuereingaben in der 3DA-Studie

Spearman's ρ in %	MMMT	HMMT	HMHT-SB	HMHT-WL
$ \rho \approx 0$	0.00	47.06	53.57	45.45
$0.2 < \rho \leq 0.4$	0.00	35.29	42.86	45.45
$ \rho > 0.4$	100.00	17.65	3.57	9.09

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

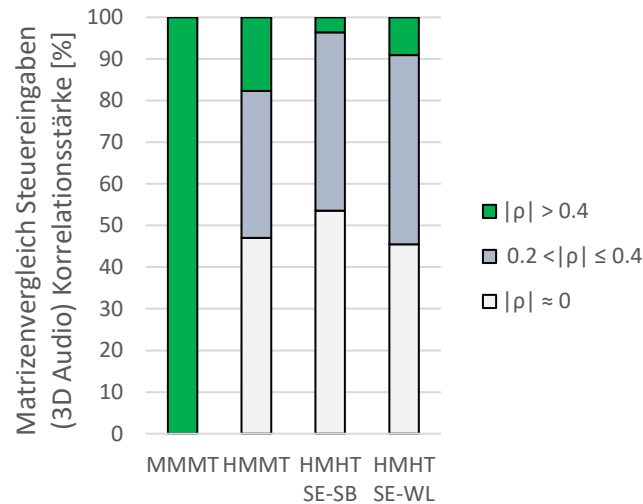


Abbildung 7.59 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Steuerereingaben in der 3DA-Studie

Neben den bislang verwendeten Simulationsmaßen wurden in der 3DA-Studie zudem Messwerte zur Bestimmung der Genauigkeit verwendet, mit denen erfasst werden sollte, wie oft der Pilot eine bestimmte Raumlage erzielte und das Flugzeug in diesem fest definierten Bereich stabilisieren konnte. Beim Vergleich dieser Matrizen der 3DA-Studie (vgl. Tabelle 7.59 und Abbildung 7.60) ergibt sich, dass die verwendeten Maße sehr heterogen sind, da die MMMT-Matrix nur 10 % an bedeutsamen Korrelationen aufweist. Diese Herangehensweise zur Messung der Nutzerleistung zeigt zudem einen äußerst geringen Anteil an bedeutsamen Korrelationen in der HMMT-Matrix (2.58 %) und ebenso niedrige Anteile in Kombination mit den Methoden zur Erfassung des Situationsbewusstseins (2.86 %) und der Workload (2.73 %).

Tabelle 7.59 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Genauigkeitsmessung Azimut in der 3DA-Studie

Spearman's ρ in %	MMMT	HMMT	HMHT-SB	HMHT-WL
$ \rho \approx 0$	70.00	74.84	70.71	69.09
$0.2 < \rho \leq 0.4$	20.00	22.58	26.43	28.18
$ \rho > 0.4$	10.00	2.58	2.86	2.73

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

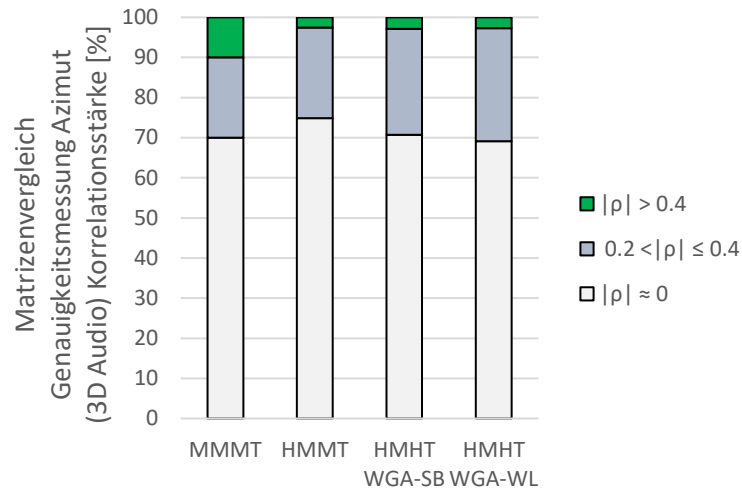


Abbildung 7.60 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Genauigkeitsmessung Azimut in der 3DA-Studie

7.2.3.2 Post Study System Usability Scale (PSSUQ)

Der PSSUQ wurde als subjektiver Fragebogen zur Erhebung der Nutzerzufriedenheit in der Machbarkeitsstudie sowie in der Untersuchung des Re-Designs einer Navigationshilfe angewendet. Der hohe Anteil an bedeutsamen Korrelationen in beiden Studien innerhalb der MMMT-Matrizen (vgl. Tabelle 7.60 und Abbildung 7.61) verdeutlicht, dass sowohl in der MS-Studie (96.32 %) als auch in der RD-Evaluierung (68.42 %) ein hoher Zusammenhang zwischen den verwendeten Mess-Items besteht.

Tabelle 7.60 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des PSSUQ

Spearman's ρ in %	MS	RD
$ \rho \approx 0$	0.00	12.11
$0.2 < \rho \leq 0.4$	3.68	19.47
$ \rho > 0.4$	96.32	68.42

Tabelle 7.61 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des PSSUQ

	MS	RD
$ \rho \approx 0$	40.36	37.60
$0.2 < \rho \leq 0.4$	31.43	41.80
$ \rho > 0.4$	28.21	20.60

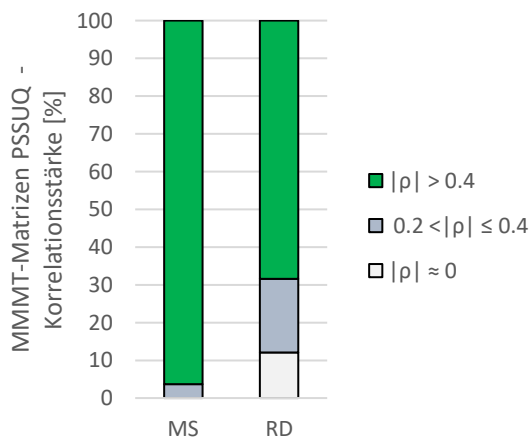


Abbildung 7.61 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des PSSUQ

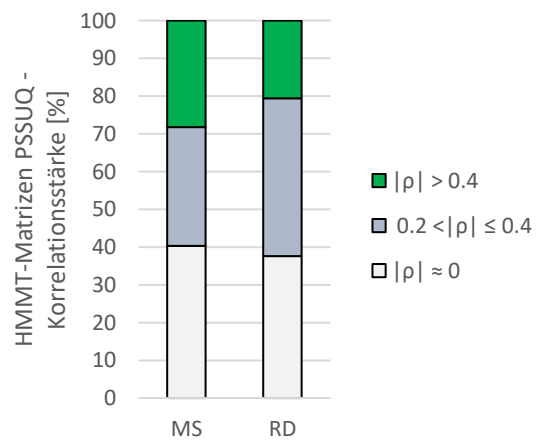


Abbildung 7.62 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des PSSUQ

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Werden die HMMT-Matrizen (vgl. Tabelle 7.61 und Abbildung 7.62) mit den HMHT-Matrizen (vgl. Tabelle 7.62 und Abbildung 7.63 sowie Tabelle 7.63 und Abbildung 7.64) verglichen, zeigt sich in beiden Fällen, dass der PSSUQ Gebrauchstauglichkeit von den Methoden zur Erfassung des Situationsbewusstseins differenzieren kann. Die HMMT-Matrizen der Machbarkeitsstudie mit 28.21 % und der RD-Studie mit 20.60 % weisen folglich höhere Anteile an bedeutsamen Korrelationen auf als die HMHT-Matrizen bezüglich der Korrelationen mit den Situationsbewusstseinsmethoden (MS = 23.57 %, RD = 12.65 %). Weniger eindeutig ist die Gegenüberstellung mit den Methoden der Workload. Nur in der Untersuchung des Re-Designs weist die HMMT-Matrix (20.60 %) einen höheren Anteil an bedeutsamen Korrelationen auf als in der HMHT-Matrix in Kombination mit Workload (16.52 %). In der Machbarkeitsstudie hingegen werden durch den PSSUQ die Methoden der Workload (50.36 %) besser beschrieben.

Tabelle 7.62 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD
$ \rho \approx 0$	52,86	50,29
$0.2 < \rho \leq 0.4$	23,57	37,06
$ \rho > 0.4$	23,57	12,65

Tabelle 7.63 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Workload

	MS	RD
$ \rho \approx 0$	22,86	49,57
$0.2 < \rho \leq 0.4$	26,79	33,91
$ \rho > 0.4$	50,36	16,52

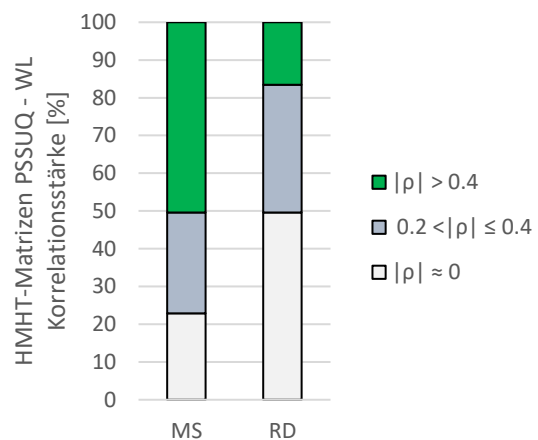
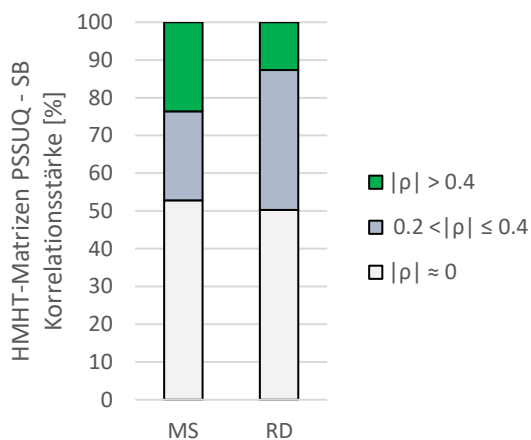


Abbildung 7.63 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Situationsbewusstsein

Abbildung 7.64 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Workload

7.2.3.3 Eye-Tracking

Als weiteres Maß zur Bestimmung der Gebrauchstauglichkeit wurden zeitabhängige Blickmaße erfasst wie beispielsweise die Blickrate pro Sekunde oder die Gesamtblickdauer während der Verwendung der zu untersuchenden Funktionen. Bei Analyse der MMT-Matrizen (vgl. Tabelle 7.64 und Abbildung 7.65) zeigt sich, dass in der MS-Studie (50 %) und in der RD-Untersuchung (47.62 %) ein erhöhter Anteil an bedeutsamen Korrelationen vorzufinden ist. Ein weniger hoher Anteil an bedeutsamen Korrelationen liegt in der SEN-Evaluierung (38.10 %) vor. Am niedrigsten jedoch ist der Anteil

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

an bedeutsamen Korrelationen in der Studie zur Untersuchung des räumlich-auditiven Warnsignals mit 33.33 %.

Tabelle 7.64 Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking Zeitleistung

Spearman`s ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	0.00	33.33	33.33	28.57
$0.2 < \rho \leq 0.4$	50.00	19.05	33.33	33.33
$ \rho > 0.4$	50.00	47.62	33.33	38.10

Tabelle 7.65 Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking Zeitleistung

	MS	RD	3DA	SEN
	33.33	39.10	54.55	30.08
	37.50	37.22	36.36	30.08
	29.17	23.68	9.09	39.85

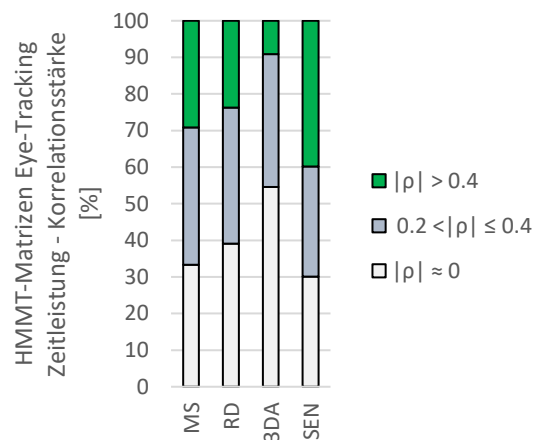
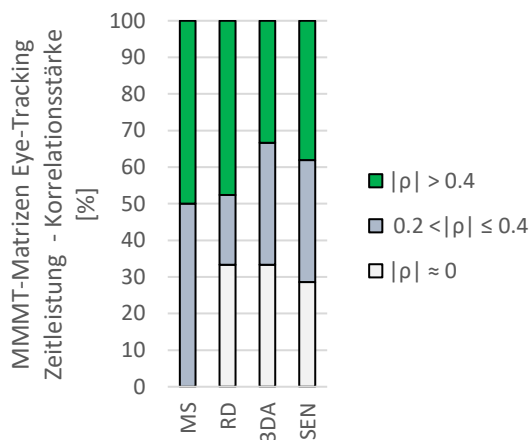


Abbildung 7.65 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking Zeitleistung

Abbildung 7.66 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking Zeitleistung

Betrachtet man die HMMT-Matrizen (vgl. Tabelle 7.65 und Abbildung 7.66) der zeitabhängigen Eye-Tracking-Maße, wird deutlich, dass die Anteile an bedeutsamen Korrelationen mit Ausnahme der 3DA-Studie (9.09 %) vergleichsweise hoch sind. Besonders hoch ist der Anteil an bedeutsamen Korrelationen in der SEN-Studie mit 39.85 %. In den Untersuchungen zur Bedienung einer Navigationshilfe sind diese Anteile im Vergleich hierzu eher moderat (MS = 29.17 %, RD = 23.68 %).

Eine Gegenüberstellung der Matrizen der konvergenten und der diskriminanten Validität ergibt, dass alle Evaluierungen niedrigere Anteile an bedeutsamen Korrelationen in den HMMT-Matrizen aufweisen als in den HMHT-Matrizen in Kombination mit Situationsbewusstsein (vgl. Tabelle 7.66 und Abbildung 7.67). Der Vergleich der HMMT-Matrizen mit den HMHT-Matrizen der Workload-Methoden (vgl. Tabelle 7.67 und Abbildung 7.68) ist hingegen weniger eindeutig, da hier zumindest in der SEN-Studie mit 32.54 % und in der RD-Evaluierung (15.53 %) die Anteile an bedeutenden Korrelationen in den HMHT-Matrizen geringer sind. Im Gegensatz hierzu deuten die Ergebnisse HMHT-Matrizen der Machbarkeitsstudie (55.36 %) und der 3DA-Studie (16.67 %) darauf hin, dass eine engere Verbundenheit zum Konstrukt Workload besteht als zu den Methoden der Gebrauchstauglichkeit.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.66 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	39.29	45.38	53.57	29.25
$0.2 < \rho \leq 0.4$	25.00	26.47	28.57	27.89
$ \rho > 0.4$	35.71	28.15	17.86	42.86

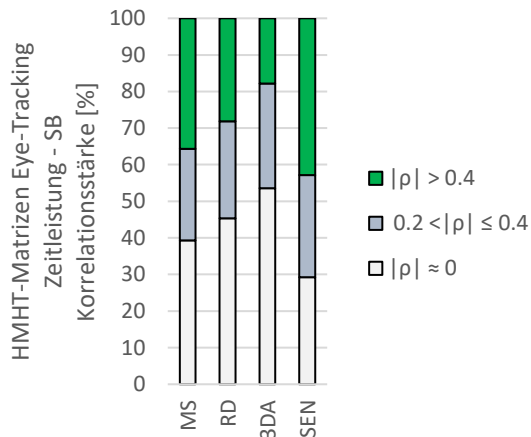


Abbildung 7.67 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Situationsbewusstsein

Tabelle 7.67 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Workload

	MS	RD	3DA	SEN
	12.50	47.83	54.55	38.89
	32.14	36.65	28.79	28.57
	55.36	15.53	16.67	32.54

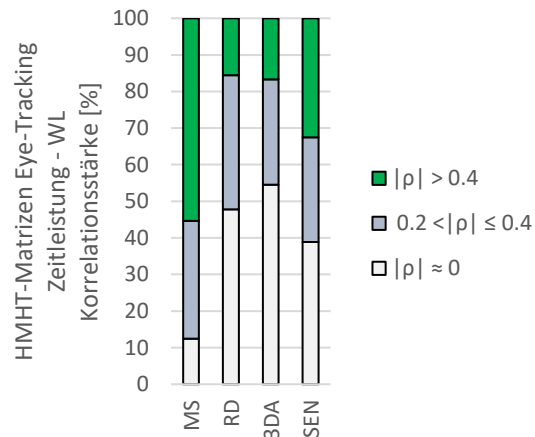


Abbildung 7.68 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Workload

Zur näheren Bestimmung der Gebrauchstauglichkeit wurde auch die Erhebung der Augenaktivität als objektives Messvorgehen genutzt, um in den Evaluierungen die Suchstrategie des Probanden sowie die Informationsdispersion zu bestimmen. Die MMT-Matrizen (vgl. Tabelle 7.68 und Abbildung 7.69) dieser Methodik weisen mit Ausnahme der SEN-Evaluierung (38.89 %) moderate Anteile an bedeutsamen Korrelationen auf. Die Mess-Items der Augenaktivität sind in den anderen Evaluierungen deutlich distinkter, da die Anteile an bedeutsamen Korrelationen gleich oder kleiner als 25 % sind.

Tabelle 7.68 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der Augenaktivität

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	40.00	38.89	41.67	36.11
$0.2 < \rho \leq 0.4$	40.00	38.89	33.33	25.00
$ \rho > 0.4$	20.00	22.22	25.00	38.89

Tabelle 7.69 Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Augenaktivität

	MS	RD	3DA	SEN
	48.81	37.96	48.56	30.72
	29.76	38.27	34.16	34.64
	21.43	23.77	17.28	34.64

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

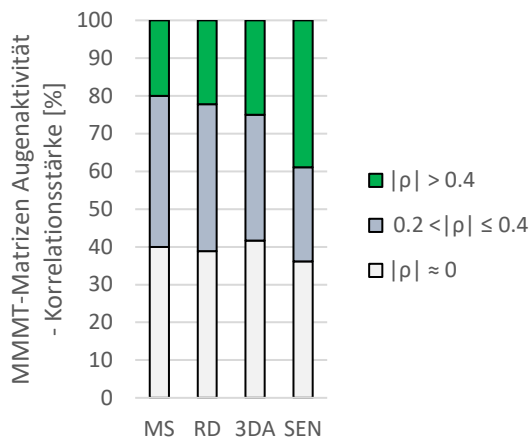


Abbildung 7.69 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Augenaktivität

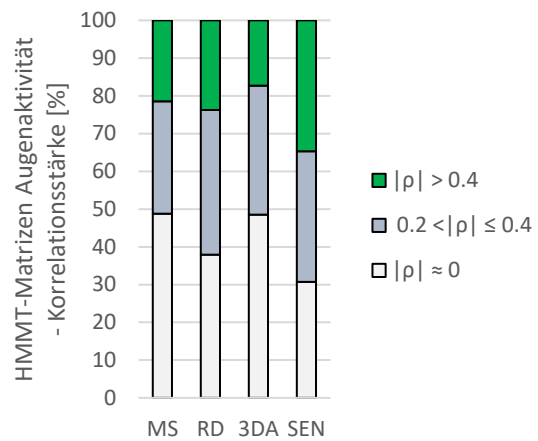


Abbildung 7.70 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Augenaktivität

Die HMMT-Matrizen (vgl. Tabelle 7.69 und Abbildung 7.70) aller Evaluierungen weisen jeweils einen vergleichsweise soliden Anteil an bedeutsamen Korrelationen auf. Insbesondere die Studie zur Untersuchung der Sensoreinstellung zeigt bezüglich der Erhebung der Augenaktivität mit 34.64 % den höchsten Anteil an bedeutsamen Korrelationen in den HMMT-Matrizen auf. Deutlich niedriger, aber immer noch moderat, sind die Anteile an bedeutsamen Korrelationen in den Studien, in denen die Einführung der Navigationshilfe (21.43 %) bzw. deren Re-Design (23.77 %) untersucht wurde. Der niedrigste Anteil an konvergenter Validität ist hingegen mit 17.28 % in der 3DA-Studie vorzufinden. Der Vergleich zu den HMMT-Matrizen (vgl. Tabelle 7.70 und Abbildung 7.71 sowie Tabelle 7.71 und Abbildung 7.72) zeigt allerdings, dass die Augenaktivität mit den Methoden zur Erfassung des Situationsbewusstseins in der MS-Studie (47.62 %) stärker korrespondiert als mit den Methoden der Gebrauchstauglichkeit. In der SEN-Studie ist der Anteil an bedeutsamen Korrelationen in der HMMT-Matrix des Situationsbewusstseins (34.92 %) nahezu gleich zu dem Anteil an konvergenter Validität mit 34.64 %. Die Erhebung von Augenaktivität war hingegen in den umfangreicheren Studien wie der RD-Studie und der 3DA-Studie besser dazu geeignet, Gebrauchstauglichkeit (RD = 23.77 %, 3DA = 17.28 %) zu beschreiben, als die Methoden zur Erfassung des Situationsbewusstseins (RD = 14.38 %, 3DA = 15.48 %). Ein Vergleich der HMMT-Matrizen mit den HMMT-Matrizen der Workload zeigt, dass die Methoden der Workload durch die gemessene Augenaktivität durchgehend besser beschrieben werden als anhand der erhobenen Maße der Gebrauchstauglichkeit, da alle Workload HMMT-Matrizen einen höheren Anteil bedeutsamer Korrelationen aufweisen.

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

Tabelle 7.70 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	3DA	SEN
$ \rho \approx 0$	26.19	59.15	51.59	37.04
$0.2 < \rho \leq 0.4$	26.19	26.47	32.94	28.04
$ \rho > 0.4$	47.62	14.38	15.48	34.92

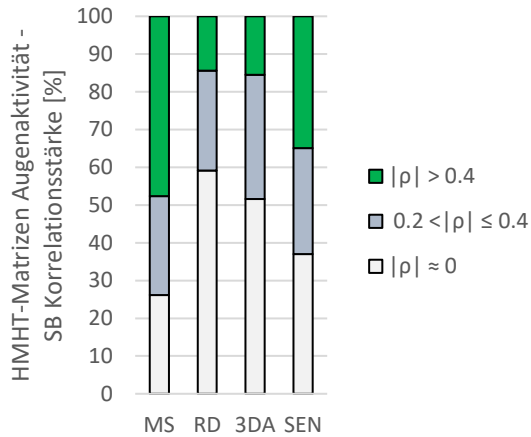


Abbildung 7.71 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Situationsbewusstsein

Tabelle 7.71 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Workload

	MS	RD	3DA	SEN
$ \rho \approx 0$	27.38	41.06	51.52	40.12
$0.2 < \rho \leq 0.4$	23.81	33.33	26.77	24.69
$ \rho > 0.4$	48.81	25.60	21.72	35.19

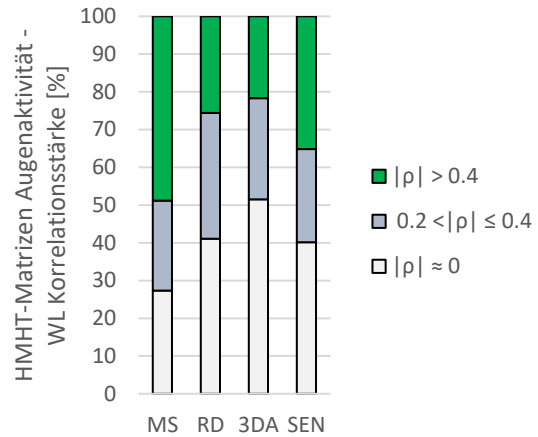


Abbildung 7.72 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Workload

Hohe Anteile an bedeutsamen Korrelationen liegen in den MMT-Matrizen (vgl. Tabelle 7.72 und Abbildung 7.73) für die aufgabenbasierte Augenaktivität vor, einem Messvorgehen, das dazu genutzt wurde, um die Augenaktivität während der Nutzung der jeweils zu untersuchenden Funktion zu erheben. Die MS-Studie weist mit 100 % den höchsten Anteil an bedeutsamen Korrelationen auf. Deutlich niedriger ist der Anteil in der SEN-Studie mit 53.57 %. Lediglich in der RD-Evaluierung ist der Anteil an bedeutsamen Korrelationen in der MMT-Matrix mit 28.57 % eher moderat.

Tabelle 7.72 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der aufgabenbasierten Augenaktivität

Spearman's ρ in %	MS	RD	SEN
$ \rho \approx 0$	0.00	57.14	21.43
$0.2 < \rho \leq 0.4$	0.00	14.29	25.00
$ \rho > 0.4$	100.00	28.57	53.57

Tabelle 7.73 Prozentuale Angabe der Korrelationsstärken in den MMT-Matrizen der aufgabenbasierten Augenaktivität

	MS	RD	SEN
$ \rho \approx 0$	43.75	36.49	36.11
$0.2 < \rho \leq 0.4$	31.25	35.14	31.25
$ \rho > 0.4$	25.00	28.38	32.64

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

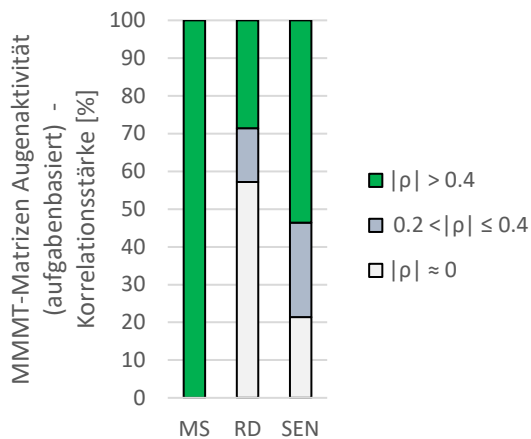


Abbildung 7.73 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Augenaktivität

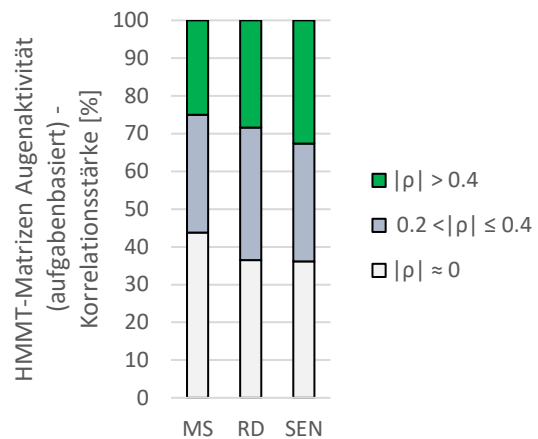


Abbildung 7.74 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Augenaktivität

Eine Gegenüberstellung der Matrizen konvergenter Validität (vgl. Tabelle 7.73 und Abbildung 7.74) mit den Matrizen diskriminanter Validität (vgl. Tabelle 7.74 und Abbildung 7.75 sowie Tabelle 7.75 und Abbildung 7.76) zeigt für die aufgabenbasierte Augenaktivität in der RD-Evaluierung einen höheren Anteil an bedeutsamen Korrelationen in der HMMT-Matrix (28.38 %) als für die HMHT-Matrix der Situationsbewusstseinsmethoden (19.85 %) und der HMHT-Matrix der Workload (17.39 %). In der Machbarkeitsstudie und der Evaluierung der Sensoreinstellung weisen die Matrizen der aufgabenbasierten Augenaktivität jedoch sowohl für die Kombination mit den Methoden zur Erfassung des Situationsbewusstseins (MS = 64.29 %, SEN = 47.62 %) als auch bezüglich der Workload-Methoden (MS = 67.86 %, SEN = 43.06 %) höhere Anteile an bedeutsamen Korrelationen auf als in Verbindung mit den Methoden der Gebrauchstauglichkeit (MS = 25 %, SEN = 32.64 %). Die Anteile an bedeutsamen Korrelationen sind dabei sowohl in den HMHT-Matrizen der Situationsbewusstseinsmethoden als auch bezüglich der HMHT-Matrizen in der Kombination mit Workload vergleichbar hoch.

Tabelle 7.74 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Situationsbewusstsein

Spearman's ρ in %	MS	RD	SEN
$ \rho \approx 0$	7.14	47.79	29.17
$0.2 < \rho \leq 0.4$	28.57	32.35	23.21
$ \rho > 0.4$	64.29	19.85	47.62

Tabelle 7.75 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Workload

	MS	RD	SEN
$ \rho \approx 0$	17.86	42.39	28.47
$0.2 < \rho \leq 0.4$	14.29	40.22	28.47
$ \rho > 0.4$	67.86	17.39	43.06

7 Ergebnisse der Validierung auf Konstrukt- und Methoden-Ebene

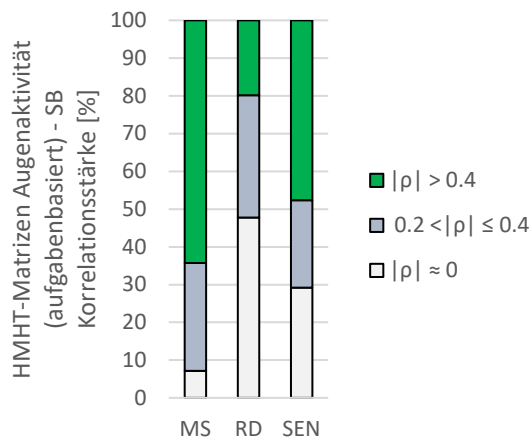


Abbildung 7.75 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Situationsbewusstsein

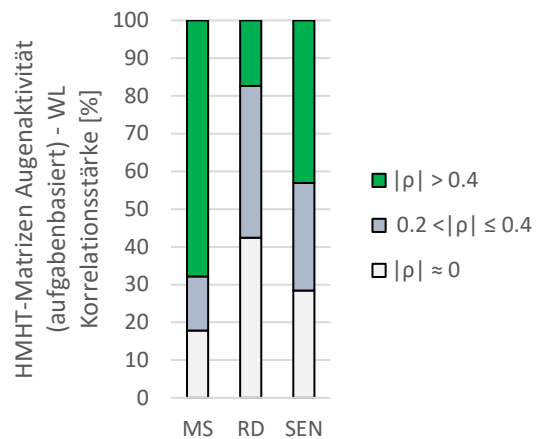


Abbildung 7.76 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Workload

7.2.4 Zusammenfassende Ergebnis-Betrachtung der verwendeten Einzelmethoden

In einer zusammenfassenden Betrachtung wird ersichtlich, dass die verwendeten Methoden anhand der eingesetzten Validierungsstrategie hinsichtlich ihrer Eignung bewertet werden können. Dies wird auch am Beispiel des PERCLOS deutlich, der als Maß hauptsächlich für die Erhebung von Ermüdung verwendet wird und somit keines der Qualitätskriterien darstellt.

Für einen Großteil der Methoden zeigen sich hohe Überschneidungen bezüglich der Qualitätskriterien, das heißt, dass die Anteile an diskriminanter Validität gegenüber den Anteilen an konvergenter Validität überwiegen. Zudem liegen jeweils hohe Korrelationen innerhalb der einzelnen Methoden und den hierin gemessenen Variablen vor. Dies gilt sowohl für die subjektiven als auch für die objektiven Messvorgehen aller Qualitätskriterien.

Das Eye-Tracking als ein objektives Maß, das in allen Konstrukten zur Erfassung der Qualitätskriterien genutzt wurde, führt mitunter zu starken Querabhängigkeiten. Des Weiteren weist die Pupillen-Dilatation als Maß zur Erfassung von Workload in den vorliegenden Evaluierungen höchstens eine Nähe zu Gebrauchstauglichkeit bei äußerst geringer konvergenter Validität auf.

Außerdem zeigt sich, dass aufgabenspezifische Methoden wie der SAGAT oder auch die funktionsrelevanten Maße zu ambivalenten Ergebnissen führen und diese nur schwer mit den globalen Methoden wie dem SART oder dem PSSUQ zu vergleichen sind.

Insgesamt ist festzustellen, dass sich für keines der Qualitätskriterien eine Methodik im Besonderen abzeichnet und eine Erfassung der Qualitätskriterien weiterhin anhand eines Methoden-Mixes vorgenommen werden sollte.

8 BEWERTUNG DER QUALITÄTSKRITERIEN UND DES MULTIDIMENSIONALEN METHODEN-INVENTARS

Nach Beschreibung der Ergebnisse auf Konstrukt-Ebene in Kapitel 7.1 und auf Methoden-Ebene in Kapitel 7.2 kann nun in Kapitel 8.1 diskutiert werden, inwiefern die verwendeten Methoden in den jeweiligen Evaluierungen dazu geeignet sind, das jeweilige Qualitätskriterium zu präzisieren. In Kapitel 8.2 wird dann die Methoden-Ebene näher beleuchtet und der Beitrag der jeweiligen Einzelmethoden hinsichtlich ihrer Eignung zur Beschreibung der Qualitätskriterien analysiert und bewertet.

8.1 Analyse und Bewertung der Qualitätskriterien auf Konstrukt-Ebene

In den folgenden Kapiteln wird diskutiert, welchen Beitrag die durchgeführten Evaluierungen und die hierin verwendeten Methoden-Kombinationen des Methoden-Inventars zur Beschreibung und Abgrenzung der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit hinzufügen. Jedes der drei Qualitätskriterien ist mit verschiedenen Kombinationen an Methoden in den Evaluierungen der Machbarkeitsstudie zur Einführung einer Navigationshilfe, dem Re-Design einer Navigationshilfe, der Überprüfung eines räumlich präsentierten Warnsignals und der Studie zur Sensorbedienung erhoben worden. Nun ist es möglich, die Distinktheit der Konstrukte anhand dieser Evaluierungen zu beschreiben.

8.1.1 Situationsbewusstsein

Nach Analyse der MonoMethod-MonoTrait (MMMT) – Matrizen des Situationsbewusstseins – zeigen sich, mit Ausnahme der 3DA-Studie, hohe Anteile an bedeutsamen Korrelationen. Das heißt im Rückschluss, dass die erhobenen Messwerte der Methoden nicht distinkt genug sind. Obwohl die verschiedenen Messwerte verschiedene Aspekte des Situationsbewusstseins messen sollten, zeigt sich hier, dass sie stattdessen etwas Ähnliches messen. Es ist zu bemerken, dass bei Analyse der MMMT-Matrizen durchaus auch mit höheren Korrelationen zu rechnen ist, da ein Konstrukt beschrieben wird, die angewendeten Mess-Items sollten aber dennoch diskriminieren, da sie unterschiedliche Aspekte des Konstruktes Situationsbewusstsein beschreiben. Das heißt, dass der Anteil an bedeutsamen Korrelationen bei den MMMT-Matrizen hier relativ hoch ist und daher nahelegt, dass viele von der Bedeutung her ähnliche Mess-Items in die Beschreibung des Methoden-Inventars eingeflossen sind. Dies wird insbesondere in der Machbarkeitsstudie deutlich. Da bei diesem Evaluationsvorgehen jedoch lediglich verschiedene Eye-Tracking-Herangehensweisen zur Beschreibung des Situationsbewusstseins genutzt wurden, ist der hohe Anteil bedeutsamer Korrelationen darauf zurückzuführen, dass die übergeordnete Methode hier ähnliche zu Grunde liegende Variablen betrachtet. In der Studie zur Sensorbedienung, die ebenso einen sehr hohen Anteil an bedeutsamen Korrelationen aufweist, wurde zur Beschreibung des Situationsbewusstseins neben dem Eye-Tracking auch der Fragebogen der 3D-Situation Awareness Rating Technique (3D SART) verwendet. Auch hier zeigen insbesondere die Messungen des Eye-Trackings, dass diese sehr nah miteinander verwandt sind, zumal eine Zusammenfassung des gleichen Datensatzes durchgeführt wird. Da bei Verwendung des

3D SART auch der errechnete Gesamtwert des Situationsbewusstseins in die Matrix miteingeflossen ist, ist auch hier von einer Begründung des höheren Prozentsatzes an bedeutsamen Korrelationen innerhalb der Methoden auszugehen. Das deutlich geringere Ausmaß an bedeutsamen Korrelationen in der 3DA-Studie spricht hingegen dafür, dass hier deutlich distinktere Mess-Items verwendet wurden als vergleichsweise in der Studie des Re-Designs. Eine Erklärung dieses Unterschiedes ist möglicherweise auf die Verwendung der Situation Awareness Global Assessment Technique (SAGAT) zurückzuführen, da die Mess-Items im Nutzerkontext der Re-Design-Evaluierung sehr stark auf die räumliche Orientierung der Probanden ausgerichtet worden sind, während bei der 3DA-Studie insbesondere Aspekte der situativen Aufmerksamkeit abgefragt worden sind.

Betrachtet man die Methoden-Kombinationen zur Beschreibung des Situationsbewusstseins im Hinblick auf die konvergente Validität, zeigt sich in den vorliegenden Evaluierungen ein ambivalentes Bild. Insbesondere die Evaluierungen, bei denen weniger Methoden zur Beschreibung des Situationsbewusstseins genutzt wurden, zeigen einen vergleichsweise höheren Anteil an bedeutsamen Korrelationen. So ist das Ergebnis von 100 % an bedeutsamen Korrelationen in der HeteroMethod-MonoTrait (HMMT)-Matrix der Machbarkeitsstudie wenig verwunderlich, da hier zwei Herangehensweisen des Eye-Trackings verglichen wurden. Dieses Ergebnis ist somit leicht zu erklären, aber aus dem genannten Grund auch von geringerer praktischer Bedeutung. Eine Abnahme an konvergenter Validität innerhalb des Situationsbewusstseins wird daher deutlich, sobald eine weitere Methode wie der 3D SART beispielsweise im Rahmen der SEN-Studie eingeführt wird. Noch deutlicher wird dieser Einbruch in den verbleibenden Studien, in denen das Methoden-Inventar umfangreich genutzt wurde, um Situationsbewusstsein zu beschreiben. Es zeigt sich, dass die eingesetzten Methoden kaum auf eine ausreichende konvergente Validität hindeuten, um das Konstrukt Situationsbewusstsein abzugrenzen.

Dieser Eindruck bestätigt sich bei näherer Betrachtung der Ergebnisse hinsichtlich der diskriminanten Validität. Die Analyse der Ergebnisse legt die Vermutung nahe, dass beide Konstrukt-Kombinationen auf eine Konfundierung der Konstrukte hindeuten. Es lässt sich allerdings nicht eindeutig feststellen, ob die verwendeten Methoden stattdessen besser dazu geeignet wären, um Workload oder Gebrauchstauglichkeit zu beschreiben. Neben der inhaltlichen Nähe lässt sich die Konfundierung des Konstruktes Situationsbewusstsein mit dem Konstrukt Gebrauchstauglichkeit allerdings auch anhand der Verwendung der Eye-Tracking-Maße zumindest teilweise begründen. So sind Fixationen im Methoden-Inventar dazu verwendet worden, um das Situationsbewusstsein zu präzisieren, während die Sakkaden als Mess-Item für Gebrauchstauglichkeit eingesetzt worden sind. Beide Maße sind jedoch zeitlich und inhaltlich voneinander abhängig.

Zusammenfassend wird jedoch deutlich, dass die verwendeten Methoden das Konstrukt Situationsbewusstsein nicht ausreichend von den Konstrukten Workload und Gebrauchstauglichkeit abgrenzen. Dies lässt vermuten, dass das Konstrukt Situationsbewusstsein nicht distinkt genug definiert ist bzw. dass die verwendeten Methoden auch Aspekte von Workload und Gebrauchstauglichkeit messen. Dieses Ergebnis wird umso augenscheinlicher, je mehr Methoden in den Evaluierungen verwendet worden sind, um das Konstrukt Situationsbewusstsein zu erheben.

8.1.2 Workload

Die Eingrenzung des Konstruktes Workload ist mit dem kleinsten Set an Methoden vorgenommen worden. In den Evaluierungen der Machbarkeitsstudie und der Sensorbedienung wurden als methodische Vorgehen jeweils nur die methodischen Herangehensweisen des NASA Task Load Index (TLX) und des Eye-Trackings verwendet, während in der 3DA- und der RD-Studie zusätzlich die taktile Detection Response Task (TDRT) als Methode genutzt worden ist.

Bei näherer Analyse der MMT-Matrizen zeigen sich auch hier recht hohe Werte an bedeutsamen Korrelationen für die meisten Evaluierungen. Allerdings schwanken die Werte zwischen 50 % und 85 % an bedeutsamen Anteilen. Generell betrachtet sind die hohen Werte als Hinweis zu werten, dass die verwendeten Items oftmals etwas Ähnliches messen und insofern das Methoden-Inventar um diese Items reduziert werden kann. Da die Ergebnisse allerdings so sehr schwanken, ist eine Reduktion des Methoden-Inventars angezeigt. Dies sollte jedoch deshalb mit Vorsicht geschehen. Erklärungen hinsichtlich der hohen Anteile an bedeutsamen Korrelationen in den MMT-Matrizen lassen insbesondere in den Methoden des NASA TLX und des Eye-Trackings finden. Der multidimensionale subjektive Fragebogen NASA TLX beinhaltet neben sechs Dimensionen auch den errechneten Gesamtwert, der ebenso in die Item-Matrix eingegangen ist. Bei Verwendung des Eye-Trackings wurden mit einem binokularen System sowohl das linke als auch das rechte Auge erhoben und weiterhin gingen auch abhängige Werte wie beispielsweise Breite und Höhe der Pupille in die Erhebung der Pupillen-Dilatation mit ein.

Betrachtet man nun die HMMT-Matrizen als zentralen Aspekt der konvergenten Validität genauer, wird für das Qualitätskriterium Workload offensichtlich, dass die Anteile an bedeutsamen Korrelationen hinsichtlich der Workload-Methoden eher gering sind. Da dies insbesondere die 3DA-Studie sowie die RD-Evaluierung betrifft, ist dieses Ergebnis wohl sowohl auf die breitere Datenbasis als auch auf die Einführung TDRT als weitere Methode zurückzuführen. Das lässt wiederum die Vermutung zu, dass die Methoden für die Beschreibung von Workload nicht für alle Evaluierungen gleich gut geeignet waren. Da die Anteile an bedeutsamen Korrelationen in den kleineren Evaluierungen zur Sensorbedienung und in der Machbarkeitsstudie deutlich höher sind, stellt sich die Frage, ob dies auf die Verwendung von nur zwei methodischen Herangehensweisen zurückzuführen ist oder in der kleineren Datenbasis begründet liegt.

Bezüglich der Abgrenzung der Konstrukte mit Hilfe der HeteroMethod-HeteroTrait (HMHT)-Matrizen wird allerdings offenkundig, dass sich auch hier die bedeutsamen Korrelationen im gleichen Größenbereich befinden wie die HMMT-Matrizen. Die Relation bleibt daher für alle Studien erhalten. Die Studie zur Sensorbedienung zeigt, dass die Workload-Methoden eher dazu geeignet gewesen wären, um Situationsbewusstsein zu beschreiben. Da in dieser Studie zur Erhebung des Situationsbewusstseins allerdings lediglich Eye-Tracking und der 3D SART verwendet wurde, lassen sich die hohen Anteile an bedeutsamen Korrelationen auch durch die Nähe des 3D SART mit dem NASA TLX erklären. So weist eine Überprüfung des 3D SARTs häufig auf Korrelationen mit Leistung und Workload auf (s.a. Selcon et al., 1991; Stanton et al., 2011). Letzteres womöglich auch deshalb, weil so-

wohl der SART als auch der NASA TLX sensitiv auf eine Veränderung der Aufgaben-Schwierigkeit reagieren (Selcon et al., 1991). Beide subjektiven Methoden erfassen aber u. a. die Beanspruchung kognitiver Ressourcen (s.a. Hart & Staveland, 1988; Taylor, 1990) und dieser Wert fließt jeweils in einen Gesamtwert ein, der auch in der Item-Matrix erfasst worden ist. Dies beeinflusst zwar auch die Evaluierung des Re-Designs und die 3DA-Studie, jedoch in einem geringeren Ausmaß, da hier das Methoden-Inventar hinsichtlich Situationsbewusstsein umfangreicher genutzt wurde. Mit Blick auf die Machbarkeitsstudie und das Re-Design wird bei Analyse der beiden HMHT-Matrizen allerdings sichtbar, dass die verwendeten Workload-Methoden besser geeignet waren, Gebrauchstauglichkeit zu beschreiben als Situationsbewusstsein oder das eigene Konstrukt Workload. Eine mögliche Erklärung hierfür ist, dass die Probanden die subjektiven Aussagen zum Thema Workload auch verstärkt in Richtung Gebrauchstauglichkeit interpretierten. Da aber lediglich der NASA TLX als einziges subjektives Instrument verwendet wurde, kann dies nur ein Teil der Erklärung sein, obwohl insbesondere die Subskalen ‚Ressourcen-Einsatz‘ und ‚Leistung‘ für eine Nähe zum Konstrukt Gebrauchstauglichkeit sprechen. Vielmehr ist allerdings davon auszugehen, dass die Konstrukte auch auf globaler Ebene und nicht nur im Bereich der hier untersuchten Evaluierungen eng miteinander verwoben und nicht distinkt genug sind.

8.1.3 Gebrauchstauglichkeit

Zur Bestimmung der Gebrauchstauglichkeit wurden verschiedene Aspekte der Nutzerleistung wie beispielsweise verschiedene Formen der Zeitleistung und Simulator-Daten als Flugleistung in der 3D-Studie verwendet sowie der Post Study System Usability Questionnaire (PSSUQ) zur Erfassung der Nutzerzufriedenheit und verschiedene Messwerte des Eye-Trackings betrachtet.

Die verwendeten Methoden zur Bestimmung der Gebrauchstauglichkeit scheinen innerhalb der Methoden deutlich besser zu diskriminieren als dies bei den Konstrukten Situationsbewusstsein oder Workload der Fall ist. Das heißt, dass die verwendeten Items sich gegenseitig ergänzen und die Kriterien von Gebrauchstauglichkeit wie Effektivität, Effizienz und Nutzerzufriedenheit besser voneinander abgrenzen können. Folglich könnten die Gebrauchstauglichkeitsmethoden auf weniger Mess-Items reduziert werden als dies bei den anderen beiden Konstrukten der Fall ist.

Im Vergleich zu den anderen beiden Konstrukten wird deutlich, dass das Konstrukt Gebrauchstauglichkeit die höchsten Anteile an bedeutsamen Korrelationen erzielt. Dies ist insbesondere deshalb von Bedeutung, da die Gebrauchstauglichkeit als Konstrukt auch in den Studien mit einer größeren Stichprobe und einem umfangreicheren Methoden-Set stabiler beschrieben wird und daher wohl eindeutiger formuliert ist als die Konstrukte Workload und Situationsbewusstsein.

Auch der Vergleich mit den HMHT-Matrizen zeigt, dass Gebrauchstauglichkeit insbesondere im Rahmen der 3D-Studie und der RD-Evaluierung tatsächlich innerhalb der eigenen Methoden besser erklärt wird als bei der Gegenüberstellung mit den anderen beiden Konstrukten. Allerdings zeigen sich bei der Machbarkeitsstudie und bei der Evaluierung der Sensorbedienung erneut deutliche Querabhängigkeiten der Konstrukte – jedoch auf unterschiedliche Art und Weise. Die Methoden der Gebrauchstauglichkeit wären in der Machbarkeitsstudie besser dazu geeignet gewesen, um Work-

load zu erheben, während in der Evaluierung zur Sensorbedienung die verwendeten Methoden besser zur Erhebung von Situationsbewusstsein genutzt worden wären. Diese Diskrepanz kann möglicherweise auf eine Beeinflussung durch den Evaluierungsgegenstand der jeweiligen Studien zurückzuführen sein, da in der Machbarkeitsstudie insbesondere die Arbeitserleichterung durch Einführung einer Navigationshilfe im Fokus stand, während bei der Sensorbedienung ein hoher Wert auf die Wahrnehmungsleistung der Probanden gelegt wurde.

8.1.4 Zusammenfassende Betrachtung der Konzepte

Die vorhandenen Matrizen wurden zusammengefasst und gesondert betrachtet. Der hohe Wert an bedeutsamen Korrelationen in den MMMT-Matrizen für das Situationsbewusstsein lässt sich mitunter auf den starken Einfluss der Machbarkeitsstudie zurückführen, in welcher das Situationsbewusstsein mit Methoden der Blickerfassen gemessen wurde – andere Methoden jedoch keinen Einfluss hatten. Ein mitigierender Einfluss der verbleibenden Studien wird allerdings ebenso offensichtlich, wenn die Anteile an bedeutsamen Korrelationen zusammenfassend betrachtet werden. Auch bei der Bestimmung von Workload wurden verschiedene Messwerte des Eye-Trackings verwendet, die zu einem Wert in dieser Höhe führen könnten. Zudem wurden in allen Konstrukten auch die errechneten und zusammenfassenden Gesamtwerte verschiedener Methoden wie beispielsweise des 3D SARTs, des NASA TLX oder des PSSUQs integriert. Gerade in Bezug auf das Eye-Tracking bedeuten die aufgezeigten Ergebnisse im Rückschluss, dass das Methoden-Inventar für alle drei Konzepte reduziert werden kann, um diese effizienter zu messen. Das heißt, ein ökonomischeres Vorgehen ist möglich, ohne die vorhandenen Ergebnisse zu verzerren. Eine Reduktion der Matrix um die Gesamtwerte könnte des Weiteren zu einer distinkteren Beschreibung der Konstrukte führen.

Hinsichtlich der konvergenten Validität wird deutlich, welchen Einfluss die Machbarkeitsstudie bzgl. der Beschreibung des Situationsbewusstseins hat, da ein Ausklammern dieser Evaluierung dazu führen würde, dass dieses Konstrukt im Speziellen deutlich schlechter aufgeklärt wird als die anderen beiden Konstrukte. Unter Beibehaltung der Machbarkeitsstudie würde hingegen das Konstrukt Situationsbewusstsein deutlich höhere Anteile an konvergenter Validität aufweisen als an diskriminanter Validität und wäre somit am besten zu präzisieren und auch abzugrenzen.

Bei der Zusammenfassung der Daten verschwindet weiterhin der Eindruck, dass die Gebrauchstauglich deutlich besser als die anderen beiden Konzepte beschrieben ist, da die kleineren Evaluierungen wie die Machbarkeitsstudie und die Sensoruntersuchung hier das Ergebnis relativieren und zu einer insgesamt niedrigeren konvergenten Validität führen.

In das Konzept Gebrauchstauglichkeit sind insbesondere die Leistungsdaten der Probanden – wie beispielsweise Bearbeitungszeiten und Flugleistungsdaten – eingeflossen. Es ist daher anzunehmen, dass diese Daten besonders empfindlich auf eine Veränderung der Workload reagieren. Dies ist auch – wenn auch etwas weniger deutlich – bei der Verbindung von Workload und Situationsbewusstsein der Fall.

Eine Erklärung für die hohen Anteile an diskriminanter Validität zwischen Situationsbewusstsein und Gebrauchstauglichkeit könnte in der Ähnlichkeit der SAGAT-Werte sowie der Leistungen bezüglich der Gefahren-Detektierung mit den Leistungswerten der Gebrauchstauglichkeitsmethoden liegen. Hinzu kommt die Verwendung der jeweiligen Blickparameter, die stark voneinander abhängig sein sollten. Da Situationsbewusstsein u.a. mit Fixationen und Blicken und Gebrauchstauglichkeit mit Sakkaden und Transitionen gemessen wurde, ist ein Teil der Abhängigkeit im Datensatz plausibel.

Eine mögliche Erklärung für die hohe diskriminante Validität zwischen Situationsbewusstsein und Workload ist natürlich, dass auch das Situationsbewusstsein von Workload moderiert wird, aber dennoch von der eigentlichen Definition ausgenommen wird (Endsley, 1988a). Zudem zeigt sich, dass beispielsweise der 3D SART auch auf Veränderungen der Aufgaben-Schwierigkeit reagiert (Selcon et al., 1991) und somit auch die Beanspruchung der Aufmerksamkeits-Ressourcen beeinträchtigt. Die Beanspruchung der Aufmerksamkeits-Ressourcen ist eine Variable, die ebenso im Konstrukt Workload mit dem NASA TLX erhoben wird (Hart & Staveland, 1988). Auch wenn sich dieser Aspekt nur auf wenige Messwerte bezieht, sollten zumindest die verbleibenden Methoden deutlicher diskriminieren.

Des Weiteren unterscheidet sich die Erhebung des Blickverhaltens für die Konstrukte Situationsbewusstsein und Workload. Workload wird in den Evaluationsmatrizen durch andere Messwerte wie beispielsweise Pupillenerweiterung und Lidschluss erfasst, während Situationsbewusstsein mit Hilfe von Fixationen und Blicken beschrieben wird und Gebrauchstauglichkeit durch die Augenaktivität. Das heißt in Rückschluss, dass Gebrauchstauglichkeit und Situationsbewusstsein hinsichtlich des Eye-Trackings stärker verknüpft werden als diese beiden Konstrukte in dem Konstrukt Workload.

In der Gesamtschau wird daher deutlich, dass die Konzepte sehr eng miteinander verschmolzen sind und nicht als disjunkte Konstrukte gesehen werden können (s.a. Billeriß et al., 2018). Diese Aussage ist mit Sicherheit auch auf die Evaluierung im Cockpit-Simulator zurückzuführen, da eine neue Funktion im Cockpit nur dann als gebrauchstauglich angesehen werden kann, sofern sie auch dem Piloten die nötigen Informationen zur Bildung seines Situationsbewusstseins liefert und gleichzeitig sicherstellt, dass dessen Workload nicht überstrapaziert wird. Weiterhin kann ein Design auch dann nicht für den Gebrauch geeignet sein, wenn diese zu einer übermäßigen Erhöhung der Workload führt und somit die Bildung des Situationsbewusstseins beeinträchtigt. Dies kann aber nur eine Erklärung für die hohe Querabhängigkeit der Konstrukte sein. Vielmehr liefern diese Ergebnisse darüber hinaus einen deutlichen Hinweis dahingehend, dass die Konstrukte generell nicht voneinander getrennt gesehen werden können – bzw. nicht eindeutig genug definiert sind, um als unterschiedliche Konstrukte zu gelten. Inwiefern die einzelnen Methoden jeweils zu diesem Ergebnis beitragen bzw. welche Methoden besser diskriminieren, wird im Folgenden offengelegt.

8.2 Analyse und Bewertung der Einzelmethode und des Methoden-Inventars

Nach genauerer Bewertung der Qualitätskriterien unter Betrachtung der vier Evaluationen ist es notwendig, den Fokus auf das verwendete Methoden-Inventar und den Beitrag der Einzelmethode

hinsichtlich der in Kapitel 8.1 diskutierten Erkenntnisse zu legen. Auf diese Weise ist es möglich, die Methoden hinsichtlich ihrer Weiterverwendung, ihrer möglichen Anpassung oder deren potenziellen Zurückweisung zu bewerten. Eine Übersicht der in den Evaluierungen verwendeten Einzelmethoden ist in Kapitel 7.2 unter Tabelle 7.13 zu finden.

8.2.1 Methoden zur Erfassung des Situationsbewusstseins

Insbesondere bezüglich des Situationsbewusstseins wurde das Methoden-Inventar breit angelegt, da hier fünf verschiedene methodische Vorgehen und deren Items untersucht wurden. Vor allem in der RD-Evaluierung und in der 3DA-Studie wurden unterschiedliche Methoden verwendet. Allerdings enthielten diese Evaluierungen auch funktionsrelevante Methoden wie die Sortieraufgabe und die Gefahren-Detektierung, die als solches für die jeweilige Studie entwickelt oder angepasst worden sind und somit nur anhand einer einzelnen Evaluierung validiert werden konnten.

8.2.1.1 3D Situation Awareness Rating Technique (3D SART)

Bei Verwendung des 3D SARTs sind die Korrelationen innerhalb der Methode nur in der Evaluierung zur Sensorbedienung hoch und in den verbleibenden Studien deutlich geringer, jedoch immer noch moderat bis hoch. Da in die Matrizen auch der Gesamtwert zur Beschreibung des Situationsbewusstseins, der aus den drei verwendeten Mess-Items errechnet wird, eingeht, ist davon auszugehen, dass dieser Wert den hohen Anteil an bedeutsamen Korrelationen innerhalb der Methode mit beeinträchtigt hat. Eine Reduktion der Mess-Items an dieser Stelle wäre demnach bezüglich der Validierungsstrategie sinnvoll. Da jedoch der errechnete Wert Situationsbewusstsein in einem einzigen Wert beschreibt und so zur Unterscheidung der Designs beitragen kann, bedeutet dies nicht, dass er per se aus dem Methoden-Inventar entfernt werden muss.

Eine Gegenüberstellung von konvergenter und diskriminanter Validität zeigt hinsichtlich des 3D SARTs ein ambivalentes Bild. Bei genauerer Analyse im Rahmen der Re-Design-Evaluierung würde der 3D SART aus dem Methoden-Inventar entfernt werden, während der Fragebogen in Anbetracht der 3DA-Studie im Methoden-Inventar verbleiben müsste.

Des Weiteren zeigen sowohl die SEN-Evaluierung als auch die RD-Studie eine enge Verbundenheit des 3D SARTs mit dem Konstrukt Workload auf. Eine mögliche Erklärung ist hier, dass Aspekte des 3D SARTs wie beispielsweise das Mess-Item zur Beschreibung der Beanspruchung der Aufmerksamkeits-Ressourcen mit Workload-Items des NASA TLX korrelieren. Dies ist auch stimmig mit den Erkenntnissen von Selcon et al. (1991), die aufzeigen, dass der 3D SART anhand der Aufgaben-Schwierigkeit moderiert wird. Allerdings weisen die Autoren zusätzlich nach, dass der 3D SART im Gegensatz zum NASA TLX auch sensitiv auf den Erfahrungshintergrund der untersuchten Operateure reagiert.

Die Aufklärung des Situationsbewusstseins mit Hilfe des 3D SARTs erweist sich zwar nicht als besonders hoch und stabil, eine Weiterverwendung des 3D SART ist jedoch aus ökonomischen Gründen angeraten. Der schnelle und leichte Einsatz rechtfertigt dessen Verwendung und den in diesem Rahmen verbundenen Informationsgehalt. Als Einzelmethode sollte der 3D SART jedoch nicht zur Bemessung des Situationsbewusstseins verwendet werden.

8.2.1.2 *Situation Awareness Global Assessment Technique (SAGAT)*

Die im Rahmen des SAGATs verwendeten Items verhalten sich zueinander sehr unterschiedlich und sind daher auch mit anderen Methoden, die Situationsbewusstsein vorgeblich einheitlich erfragen, wie beispielsweise der 3D SART, nur schwer zu vergleichen. Dies zeichnet sich auch deutlich in den MMMT-Matrizen ab. Da in der RD-Studie das Situationsbewusstsein hauptsächlich in Form von retrospektiver räumlicher Orientierung erfragt worden ist und in der 3DA-Evaluierung auch Aspekte der Leistung sowie situative Aufmerksamkeit eingebunden worden sind, ist dieser Unterschied hierdurch erklärbar.

Es wird auch bei Gegenüberstellung des SAGATs mit den verbleibenden Methoden des Situationsbewusstseins deutlich, dass diese nur geringfügig miteinander bedeutsam korrelieren und somit nur einen kleinen Anteil an konvergenter Validität aufweisen. Allerdings zeigt sich, dass die Items, die in der 3DA-Studie verwendet worden sind, Situationsbewusstsein marginal besser von den Mess-Methoden zu Workload und der Gebrauchstauglichkeit abgrenzen als in der RD-Studie. Eine mögliche Erklärung ist, dass die verwendeten Items auch dem eigentlichen Vorgehen des SAGATs besser entsprechen als die Items der RD-Studie, in der ein ‚Einfrieren des Szenarios‘ als zu intrusiv betrachtet worden ist.

Die Schwierigkeit hinsichtlich der Analyse des SAGATs liegt in den Mess-Items selbst, da diese auch für die beiden Evaluierungen nicht vergleichbar sind. Untersucht wurden Aspekte des Situationsbewusstseins, die sich auf Grund der Versuchsaufbauten und des Nutzerkontextes deutlich unterscheiden und Situationsbewusstsein somit aufgabenspezifisch messen. Dies würde daher auch den Ergebnissen von Salmon et. al. (2007) entsprechen, wonach SART und SAGAT kaum miteinander zusammenhängen. Dieses Ergebnis ist in einer Studie zur Untersuchung in einer Command-und-Control-Aufgabe erzielt worden; nach anschließender Durchführung eines qualitativen Vergleichs der Methoden schlussfolgern die Autoren, dass der SART globale Aspekte des Situationsbewusstseins misst, während sich der SAGAT auf spezifische Aspekte bezieht (Salmon et al., 2007). Allerdings sind in beiden hier vorliegenden Evaluierungen auch aufgabenspezifische Messungen wie beispielsweise im Rahmen des Eye-Trackings oder bei Verwendung der Gefahren-Detektierung mit eingeflossen, sodass hier durchaus mit einem höheren Anteil an bedeutsamen Korrelationen bezüglich der konvergenten Validität zu rechnen war.

Eine klare Entscheidung für den Verbleib des SAGATs im Methoden-Inventar kann hier nicht getroffen werden, da weitere Untersuchungen notwendig sind, um hier konvergente Validität gegenüber diskriminanter Validität abzuwägen. Insgesamt gesehen zeigt sich jedoch die Aufklärung des Situationsbewusstseins anhand des SAGATs bei gleichzeitiger Betrachtung des Aufwandes als gering.

8.2.1.3 *Sortieraufgabe*

Eine Gegenüberstellung aller Matrizen des Multi-Trait/ Multi-Method-Vorgehens mit der Sortieraufgabe in der RD-Evaluierung lässt vermuten, dass diese funktionsrelevante Methode im Rahmen ihres Einsatzes weniger gut zur Aufklärung des Situationsbewusstseins beiträgt. Es wird offensichtlich, dass

die in der RD-Studie verwendeten Items so stark miteinander korrelieren, dass diese redundant sind und das Methoden-Inventar daher gekürzt werden sollte. Beispielsweise ist eine Unterscheidung nach idealer Handlungsabfolge und nach optimaler Szenario-Ordnung (vgl. hierzu Kapitel 5.2.5.1) in das Methoden-Inventar eingegangen, die mit Blick auf die Validierungsstrategie nicht notwendig erscheint.

Bei Gegenüberstellung der Matrizen ist ersichtlich, dass eine eindeutige Abgrenzung zu den Konstrukten Workload und Gebrauchstauglichkeit nicht vorgenommen werden kann. Stattdessen wird ein deutlicher Bezug zum Konstrukt der Gebrauchstauglichkeit sichtbar. Im Rückschluss sollte daher eine Verwendung der Sortieraufgabe für das Konstrukt Gebrauchstauglichkeit angestrebt werden, da dieses Vorgehen anscheinend eher im Sinne einer effektiven und effizienten Handlungsumsetzung als der Handlungsplanung zu interpretieren ist.

8.2.1.4 Gefahren-Detektierung

Die Gefahren-Detektierung ist als Wahrnehmungsaufgabe im Zuge der 3DA-Untersuchung in das Methoden-Inventar integriert und als funktionsrelevante, kontextspezifische Aufgabe eingesetzt worden. Nach Analyse und Vergleich der Matrizen wird ersichtlich, dass die verwendeten Items zwar ausreichend untereinander differenzieren, aber auch eine Kürzung angezeigt ist. Da innerhalb dieser Methodik neben dem objektiven Leistungswert der gemittelten Entdeckungsrate auch Werte an Eigen- und Fremdbewertungen sowie die daraus ermittelte Abweichung eingeflossen sind, kann die Ähnlichkeit der eingesetzten Messwerte erklärt werden.

Im Vergleich zur konvergenten Validität ist die diskriminante Validität hinsichtlich der Konstrukte Workload und Gebrauchstauglichkeit deutlich größer. Im Umkehrschluss wäre daher diese Methode zur Ermittlung von Workload besser geeignet gewesen als in ihrer aktuellen Verwendung für das Konstrukt Situationsbewusstsein. Eine mögliche Erklärung ist hier, dass der im Rahmen dieser Methodik erhobene Leistungsparameter mit den Leistungswerten der Konstrukte Workload (beispielsweise Parameter der TDRT) und der Gebrauchstauglichkeit (beispielsweise der Flugleistungswerte) verwandt ist und anhand der Bedingungen des Versuchsaufbaus ähnlich beeinflusst wurde.

Da es sich bei der Gefahren-Detektierung um eine anwendungsspezifische Methodik handelt, ist eine Anpassung der Wahrnehmungsaufgabe an den jeweiligen Kontext notwendig. Hinsichtlich der Konstrukt-Zuweisung wäre diese Aufgabe jedoch auf Grund der vorliegenden Ergebnisse eher der Bestimmung von Workload oder der Gebrauchstauglichkeit dienlich, obwohl sie vom theoretischen Ansatz her eindeutig dem Situationsbewusstsein zuzuordnen ist.

8.2.1.5 Eye-Tracking

Zur Erhebung der Aufmerksamkeit mittels Eye-Tracking wurden die Daten sowohl global wie auch aufgabenspezifisch erfasst. Für beide Herangehensweisen ergeben sich innerhalb der Methode starke Schwankungen bezüglich der Anteile an bedeutenden Korrelationen, die sich zum einen darauf zurückführen lassen, dass in der Machbarkeitsstudie ein monokularer Eye-Tracker und in den anderen beiden Studien ein binokulares System verwendet wurde, mit dem daher ein größerer Datensatz

aufgezeichnet wurde (Billeriß et al., 2018). Ebenso zeigt sich für die globale Methodik, dass in den Studien mit breiterer Datenbasis der Anteil an bedeutsamen Korrelationen geringer ist. Aufgrund der stark schwankenden Ergebnisse ist eine Reduktion der Mess-Items nur mit Vorsicht vorzunehmen und sollte sich auf ähnliche Werte beschränken wie beispielsweise die Anzahl der Fixationen für das linke oder das rechte Auge.

Da in der Machbarkeitsstudie Aufmerksamkeit mit Eye-Tracking-Maßen gemessen wurde, die mit Bezug auf das ganze Szenario erhoben wurden und mit den Eye-Tracking-Werten, die während der Bedienung der Navigationshilfe erfasst wurden, verglichen wurde, muss hier von einer weitergehenden Interpretation der konvergenten Validität abgesehen werden.

Die verbleibenden Studien zeigen hierzu deutlich niedrigere Anteile an konvergenter Validität, die sich im Vergleich mit der diskriminanten Validität jedoch als ambivalent erweisen. Die globale Eye-Tracking-Herangehensweise zur Erfassung von Aufmerksamkeit differenziert gegenüber Workload nicht, aber gegenüber Gebrauchstauglichkeit. Die Verknüpfung von der mit Eye-Tracking erhobenen, visuellen Aufmerksamkeit mit dem Konstrukt Gebrauchstauglichkeit lässt sich damit erklären, dass auch in diesem Konstrukt Eye-Tracking-Maße wie beispielsweise die Erhebung von Augenaktivität eingeflossen sind. Da hierunter Messwerte wie Sakkaden, Transitionen und deren Dauer fallen, könnte diese Beziehung hierin begründet sein.

Bei Betrachtung der aufgabenspezifischen Eye-Tracking-Herangehensweise wird hingegen nicht Gebrauchstauglichkeit, sondern Workload besser beschrieben als das Situationsbewusstsein. Somit wird offensichtlich, dass mittels Eye-Tracking Situationsbewusstsein nicht von den Konstrukten Workload und Gebrauchstauglichkeit abgegrenzt werden kann. Beide Eye-Tracking-Herangehensweisen zeigen also, dass sie weniger gut geeignet sind, um Situationsbewusstsein zu erfassen – insbesondere, wenn unterschiedliche Methoden zur Erfassung dieses Konstrukts herangezogen werden. Die Frage, ob dieses Messvorgehen stattdessen eher der Workload- oder den Gebrauchstauglichkeitsmethoden zugeordnet werden kann, kann hingegen nicht beantwortet werden, da hier die Ergebnisse uneindeutig sind. Allerdings ist unter Betrachtung der Ergebnisse eine Umstrukturierung der Matrix indiziert bzw. ein Exkludieren der Werte aus den Evaluierungsmatrizen, die unmittelbar zusammenhängen wie beispielsweise Fixationen und Sakkaden oder Blicke und Transitionen.

8.2.2 Methoden zur Erfassung der Workload

Um Workload zu präzisieren und von den anderen Qualitätsmerkmalen abzugrenzen, sind in den Evaluierungen drei unterschiedliche Methoden – der NASA TLX, die TDRT und Eye-Tracking-Maße – verwendet worden. Insgesamt ist im Hinblick auf das Methoden-Inventar zur Bestimmung von Workload die geringste Anzahl an verschiedenen Methoden verwendet worden. Da jedoch auch unterschiedliche Herangehensweisen des Eye-Trackings betrachtet werden, kann insbesondere diese als objektiv eingestufte Messmethode hinsichtlich der Erhebung von Workload validiert und diskutiert werden.

8.2.2.1 NASA Task Load Index (NASA TLX)

Auch in der Anwendung des NASA TLX, als weit verbreitetem Verfahren zur Erhebung von Workload, fällt auf, dass die Mess-Items innerhalb der Methode moderat bis hoch miteinander korrelieren. Ein Hinweis darauf, dass die gemessenen Items durchaus differenzieren, ist in den Studien zur Untersuchung eines räumlich-auditiven Warnsignals und in der Evaluierung der Sensorbedienung zu finden. Allerdings spricht das Ergebnis dafür, dass eine so differenzierte Unterscheidung nach sechs Dimensionen gar nicht notwendig ist. Für eine Zusammenfassung der Werte sprechen auch Ergebnisse von Hendy, Hamilton & Landry (1993), die herausgefunden haben, dass eine eindimensionale Skala sensibler ist, um subjektiven Workload hinsichtlich der Anforderungen an die Informationsverarbeitung des Nutzers zu erheben als es der NASA TLX ist. Dieses Ergebnis wird von Hill et. al. (1992) insofern gestützt, dass diese nach einem Vergleich von zwei multidimensionalen und zwei eindimensionalen Erhebungsverfahren empfehlen, eine eindimensionale Skala wie die Overall Workload Scale (Vidulich & Tsang, 1987) zur Identifikation von Extrempunkten zu nutzen, während der NASA TLX als Diagnoseinstrument hinsichtlich seiner Subskalen für weitere Befragungen oder Maßnahmen verwendet werden kann. Zumindest könnte aber der zusammenfassende Workload-Wert aus dem Methoden-Inventar exkludiert werden, um die Anteile an bedeutsamen Korrelationen in den MMMT-Matrizen zu konsolidieren.

Der Rückgang an konvergenter Validität in der RD-Evaluierung und der 3DA-Studie ist mit der Verwendung der TDRT als weitere Methode zur Erfassung der Workload unter objektiven Kriterien zu begründen. Im Rückschluss heißt das, dass der NASA TLX wohl weniger stark mit den gemessenen Items der TDRT korrespondiert. Stellt man die konvergente Validität der diskriminanten Validität gegenüber, so wird deutlich, dass der NASA TLX sowohl in der 3DA-Evaluierung als auch in der Untersuchung zur Sensorbedienung die Konstrukte voneinander abgrenzen kann und hierbei das eigene Konstrukt Workload besser beschreibt.

Ein Verbleib des NASA TLX im Methoden-Inventar ist somit auf Grund dieser Ergebnisse sinnvoll, da er eine Differenzierung der Konstrukte zulässt, bekannt sowie leicht verständlich ist und schnell angewendet werden kann. Lediglich in der Abgrenzung von Situationsbewusstsein scheint dieses Vorgehen weniger gut geeignet. Zumindest lassen die Ergebnisse eine engere Verbindung zum Konstrukt Situationsbewusstsein vermuten als zur Gebrauchstauglichkeit. Begründet kann dies in der Verwendung des 3D SART sein, der auch die Beanspruchung von Aufmerksamkeits-Ressourcen erfasst und laut Selcon, Taylor und Francis (1991) zumindest durch die Aufgaben-Schwierigkeit moderiert wird. Offensichtlich ist jedenfalls, dass in allen drei Studien, in welchen der 3D SART erhoben worden ist, der Anteil an bedeutsamen Korrelationen für das Konstrukt Situationsbewusstsein höher ist als für das Konstrukt Gebrauchstauglichkeit. Dennoch ist auch in der Machbarkeitsstudie und der RD-Untersuchung eine Nähe zu diesem Konstrukt nachweisbar, was wiederum daran liegen könnte, dass der NASA TLX eine subjektive Einschätzung der Probandenleistung erfragt, während in das Konstrukt Gebrauchstauglichkeit die objektiven Leistungsmessungen eingegangen sind.

8.2.2.2 *Tactile Detection Response Task (TDRT)*

Die TDRT wurde als weitere Maßnahme zur objektiven Erhebung von Nutzer-Workload in der RD-Untersuchung und der 3DA-Evaluierung eingeführt. Eine nähere Betrachtung der MMMT-Matrizen ergibt, dass weitere Untersuchungen mit dieser Methodik notwendig sind, bevor die Mess-Items reduziert werden können, da beide Untersuchungen durch einen sehr unterschiedlichen Anteil an bedeutsamen Korrelationen gekennzeichnet sind.

Bei Vergleich der Validitätsformen ergibt sich eine größere Nähe zum Konstrukt Gebrauchstauglichkeit als zum Konstrukt Workload. Da in die TDRT ebenso Reaktionszeiten einfließen, wie in das Konstrukt Gebrauchstauglichkeit unter dem Aspekt Nutzerleistung, bleibt zu vermuten, dass diese Probanden-Reaktionen in den Evaluierungen ähnlich beeinflusst wurden.

Da der Anteil an konvergenter Validität des Weiteren hinsichtlich der TDRT so gering ausfällt, ist diese Methodik in der angewendeten Form nicht dazu geeignet, um Workload zu erheben bzw. ist sie von den anderen im Methoden-Inventar verwendeten Workload-Methoden grundlegend verschieden. Eine Anpassung dieser Methodik wäre daher sinnvoll und notwendig. Dies könnte in Form einer genaueren Anpassung an die Norm ISO 17488:2016 geschehen, die zum Zeitpunkt der Versuchsdurchführung noch nicht zur Verfügung stand und zudem für den Automobilkontext entwickelt wurde. Eine weitere Überlegung wäre es, diese Methodik hinsichtlich der Signalentdeckungstheorie (s.a. Goldstein, 1999) auszulegen, da in diesem Fall ein Zweitreiz als Distraktor-Reiz eingeführt werden würde und somit die Antwort-Tendenzen der Probanden zur Bestimmung von Workload mitberücksichtigt werden können.

8.2.2.3 *Eye-Tracking*

Zur Erhebung von Workload sind unterschiedliche Parameter des Eye-Trackings wie die Pupillenerweiterung, Maße zur Aufmerksamkeitsfixierung und als Abgrenzung zum Konstrukt auch der PERCLOS als Indikator für Müdigkeit erhoben worden.

Da die erhobenen Werte zur Erhebung der Pupillen-Dilatation in allen vier Evaluierungen jeweils zu 100 % bedeutsam korrelieren, ist eine Reduktion der Messwerte erforderlich. Eine gesonderte Betrachtung aller Werte scheint obsolet, weil hier sowohl Werte zur Beschreibung der Pupillenhöhe und -weite sowie die Pupillenfläche eingeflossen sind. Zudem wurde in den Studien zur Evaluierung des Re-Designs, des räumlich-auditiven Warnsignals und in der Untersuchung der Sensoreinstellung ein binokulares System verwendet, welches die Pupillendaten für das linke und das rechte Auge separat erfasst. Aufgrund des vorliegenden Ergebnisses sollten daher die Messwerte auf ein Auge sowie die Pupillenfläche reduziert werden.

Ein Vergleich der konvergenten Validität mit der diskriminanten Validität verdeutlicht, dass eine Präzisierung und Abgrenzung des Konstrukts Workload anhand der Pupillenerweiterung nicht möglich ist, da nur in der Untersuchung zur Sensorbedienung der Anteil an konvergenter Validität größer ist als die diskriminante Validität bezüglich der Methoden-Sets für Situationsbewusstsein und Gebrauchstauglichkeit. Dass die Pupillenerweiterung Situationsbewusstsein in der Machbarkeits-Studie

besser erklärt als Workload, ist mit der Tatsache zu erklären, dass in dieser Studie Situationsbewusstsein allein mittels Eye-Tracking erhoben wurde. Da Workload in der Machbarkeitsstudie ergänzend mit dem NASA TLX gemessen wurde, ist daher der kleinere Anteil an bedeutsamen Korrelationen hier nachvollziehbar. Da jedoch für das Situationsbewusstsein als auch bezüglich der Pupillen-Dilatation grundsätzlich verschiedene Parameter genutzt wurden, sollten die Items im Einzelnen auf diese Verknüpfung untersucht werden.

Auf Grund der Ambivalenz der Ergebnisse, die keine Systematik erkennen lassen, sollte von einer weiteren Verwendung der Pupillenerweiterung zur Erfassung von Workload abgesehen werden.

Auch bei der Bestimmung des PERCLOS zur Erhebung und Abgrenzung von Ermüdung in Folge von Workload wird eine deutliche Abhängigkeit der Mess-Items innerhalb der Methode offensichtlich. Eine Reduktion der Mess-Items ist daher aus ökonomischen Gründen erforderlich. Auch innerhalb dieser Methodik wurden anhand des binokularen Systems für die Studie zur Erhebung des Re-Designs, der 3DA-Studie sowie der SEN-Evaluation Daten für beide Augen und ein Durchschnittswert erfasst. Eine Reduktion auf den PERCLOS-Durchschnitt, der für ein Auge erhoben wird, ist mit Blick auf die MMT-Ergebnisse daher ausreichend.

Eine Weiterverwendung dieser Methodik kann im Hinblick dieser Ergebnisse nicht empfohlen werden, da keine eindeutige Zuordnung zu einem der Konzepte erfolgen kann. Ermüdung ist eher als Anzeichen und Folge von mentaler Arbeitsbelastung dem Konstrukt Workload zuzuordnen (s.a. Hart & Wickens, 1990). Da die MMT-Matrizen aber durchweg niedrigere Korrelationen aufweisen als die HMHT-Matrizen, ist eine direkte Zuordnung zu diesem Konstrukt nicht möglich. Auch eine systematische Verbindung zu einem der anderen beiden Konstrukte ist mit der vorhandenen Datenlage nicht nachweisbar. Die Erfassung von Ermüdung ist daher deutlich von der Erhebung der Workload abzugrenzen und somit für die Erhebung dieses Konstruktes irrelevant. Es bestätigt sich somit, dass PERCLOS vermutlich einem anderen Konstrukt wie dem der Ermüdung zuzuordnen ist und nicht Workload misst.

Als weitere Eye-Tracking-Herangehensweise ist die Aufmerksamkeitsfixierung zur Präzisierung von Workload in das Methoden-Inventar eingebunden worden. Bei Analyse der MMT-Matrizen zeigt sich, dass diese Parameter nicht weiter reduziert werden können, da sich die Daten zueinander nicht kongruent verhalten. Des Weiteren führt der Vergleich der MMT-Matrizen mit den HMHT-Matrizen zu der Erkenntnis, dass das Konzept Workload mittels dieses Messvorgehens nicht ausreichend von den anderen beiden Konstrukten abgegrenzt werden kann – insbesondere nicht von der Gebrauchstauglichkeit. Dies ist insofern verwunderlich, da diese Messwerte auf Grund ihrer Definition eher dem Konstrukt Situationsbewusstsein zuzuweisen wären.

Sowohl die MMT-Matrizen als auch die Gegenüberstellung der MMT mit den HMHT-Matrizen zeigen, dass sich die Daten dieser Methodik sehr unsystematisch verhalten und daher von einer Weiterverwendung dieser Messstrategie abzusehen ist.

8.2.3 Methoden zur Erfassung der Gebrauchstauglichkeit

Zur Bestimmung der Gebrauchstauglichkeit sind drei zu unterscheidende Messstrategien verwendet worden – die Nutzerleistung, der PSSUQ als subjektiver Fragebogen und das Eye-Tracking. Die Nutzerleistung dient hier allerdings als Oberbegriff für zwei verschiedene Leistungsparameter des Nutzers – der Zeitleistung und der Flugleistung. Hinter der Zeitleistung sind hier insbesondere Bediengeschwindigkeiten sowie Reaktionszeiten zu verstehen, während die Erfassung der Flugleistung vor allem Simulator-Daten aus der 3DA-Studie zur Erhebung von Genauigkeitsparametern beinhaltet.

8.2.3.1 Nutzerleistung

Zeitleistung als eine Form der Nutzerleistung ist ein wichtiges Mess-Item zur Erhebung von Gebrauchstauglichkeit, da es den Ressourcen-Einsatz wie beispielsweise Bedienzeit kennzeichnet, die für das Design von Nutzerschnittstellen und zur Optimierung von Prozessen und Verfahren genutzt werden kann (s.a. Stanton et al., 2011). Betrachtet man die Zeitleistung der Nutzer über alle Evaluierungen hinweg, ist zu erkennen, dass dieses Methoden-Inventar nur in der Machbarkeitsstudie reduziert werden kann. In den verbleibenden Studien differenzieren die verwendeten Maße moderat bis gut.

Es wird deutlich, dass Zeitleistung nicht nur im Rahmen des Konstruktes Gebrauchstauglichkeit sinnvoll erhoben werden kann, sondern auch für die Konstrukte Workload und Situationsbewusstsein relevant ist. Die hohen Anteile an bedeutsamen Korrelationen für das Konstrukt Situationsbewusstsein können dadurch erklärt werden, dass auch hinsichtlich dieser Methoden Zeitleistung beispielsweise für die Beantwortung des SAGATs als Zweitaufgabe erhoben wurde und diese Zeiten beispielsweise mit den Bedienzeiten korrespondieren. Dass die Verknüpfung zur Workload hingegen so stark ist, lässt sich dahingehend erklären, dass die Zeitleistung der Probanden in der Sekundäraufgabe offensichtlich stark durch die Schwankungen der Arbeitsbelastung beeinträchtigt wird und auch in diesem Konstrukt Zeitleistungen methodenimmanent beispielsweise im Rahmen der TDRT erfasst werden. Gerade in der 3DA-Studie wurde Zeitleistung für mehrere verschiedene Komponenten spezifisch erhoben wie beispielsweise die Zeit, bis ein bestimmtes Flugmanöver durchgeführt wurde, die Zeit während ein Flugmanöver durchgeführt wurde sowie die verbrachte Zeit in spezifischen Flugzuständen. Anscheinend hat in diesem Fall die Vielfalt der erhobenen Zeitleistungen zu deren engeren Definition im Rahmen der Gebrauchstauglichkeit beigetragen und dazu geführt, dass Gebrauchstauglichkeit in dieser einen Studie die Konstrukte differenziert.

Offensichtlich wird somit, dass Zeit als Leistungsmaß in allen Konstrukten erhoben wird und unabhängig von den Konstrukten beeinträchtigt wird. Als Maß selbst sollte es daher weiter erfasst werden – eine Unterscheidung nach Konstrukten ist aber mit der Erhebung von Zeitleistung nicht möglich.

Als weitere Herangehensweise zur Bestimmung der Nutzerleistung wurde die Flugleistung der Probanden erfasst, um die Genauigkeit der Aufgabenbearbeitung zu beschreiben. Hierzu wurden

unterschiedliche Aspekte der Flugleistung wie beispielsweise Geschwindigkeitsleistung, Winkelgenauigkeit oder Steuereingaben anhand von Simulationsdaten in der 3DA-Evaluation erhoben.

Hinsichtlich der Methode der Geschwindigkeitsleistung wird ersichtlich, dass die verwendeten Mess-Items nicht ausreichend differenzieren und nur sehr geringe Anteile an konvergenter und diskriminanter Validität aufzeigen. Grundsätzlich aber eignet sich die Einhaltung von Geschwindigkeit als eine Form der aufgabenspezifischen Erhebung dazu, Gebrauchstauglichkeit zu erfassen und sollte daher auch für weitere Evaluierungen angedacht werden – sofern sie im jeweiligen Nutzungskontext verwendet werden kann.

Eine weitere Methode zur Bestimmung der Flugleistung in der 3DA-Evaluierung war der Aspekt der Winkelgenauigkeit während eines spezifischen Flugmanövers, deren Mess-Items auch um ein paar wenige Werte reduziert werden könnten. Im Hinblick auf den Matrizenvergleich war dieses Messvorgehen zur Bestimmung der Gebrauchstauglichkeit in der 3DA-Studie gut geeignet und sollte in weiteren Studien, in denen eine ähnliche Aufgabenspezifität angezeigt ist, als Genauigkeitsleistung weiter untersucht und in diesem Zusammenhang validiert werden.

Auch die Betrachtung der Steuereingaben als Genauigkeitsmessung der 3DA-Studie würde auf Grundlage der Ergebnisse eine Reduktion des Methoden-Inventars rechtfertigen. Dies ist jedoch im vorliegenden Fall nicht möglich. Als Messwerte wurden zum einen die Steuereingaben in der Höhe und zum anderen die Steuereingaben im Azimut verwendet. Da diese Messungen sich in ihrer Definition grundsätzlich unterscheiden, ist eine Reduktion des Methoden-Inventars nicht angezeigt, da hier vielmehr ein inhaltlicher, durch die Untersuchung bedingter Zusammenhang zu vermuten ist. Im Hinblick auf den Matrizenvergleich ist zu erkennen, dass in der 3DA-Studie die Steuereingaben gut dazu genutzt werden können, um die Konstrukte und deren Methoden voneinander abzugrenzen. Eine weitere Untersuchung der Steuereingaben unter dem Konstrukt Gebrauchstauglichkeit ist daher sinnvoll, sollte jedoch auch in einem anderen hierfür geeigneten Kontext validiert werden.

Einzig die Fluggenauigkeitsleistung im Azimut scheint für den Verbleib im Methoden-Inventar als methodisches Vorgehen im Rahmen der 3DA-Evaluierung nicht geeignet zu sein. Diese Form der Genauigkeitsmessung trägt weder dazu bei, das Konstrukt Gebrauchstauglichkeit näher zu erklären noch die anderen beiden Konstrukte, da die Anteile an bedeutsamen Korrelationen zum einen äußerst gering sind und sich nicht dazu eignen, um Gebrauchstauglichkeit als Konstrukt zu beschreiben.

Grundsätzlich bleibt bei Betrachtung der Flugleistung zu erwähnen, dass die verwendeten Simulationsdaten nur in der 3DA-Evaluierung abgeleitet und evaluiert worden sind. Somit sind diese Genauigkeitswerte der Nutzerleistung aufgabenspezifisch erfasst worden und können nur im Rahmen dieser Evaluation und den hierin verwendeten Methoden interpretiert werden. Eine Übertragung dieser Erkenntnisse ist somit nicht automatisch auf weitere Evaluierungen möglich. Allerdings zeigt sich anhand der beschriebenen Erkenntnisse, dass es Sinn ergibt, den Aufgabenkontext im Hinblick auf die Gebrauchstauglichkeit der Designänderung genau zu erfassen, da dies eine Abgrenzung dieses Konstrukts von den Konstrukten Situationsbewusstsein und Workload erlaubt. Für andere Evaluierungen sollten daher durchaus die Nutzerleistungen anhand der Simulationsdaten im Rahmen der

Gebrauchstauglichkeit erhoben werden, sofern diese auch als Maße für Effektivität und Effizienz in dieser Mensch-Maschine-Interaktion zu verstehen sind.

8.2.3.2 *Post Study System Usability Scale (PSSUQ)*

Zur Erhebung von subjektiv empfundener Nutzerzufriedenheit wurde der PSSUQ eingesetzt. Die Betrachtung der MMMT-Matrizen lässt darauf schließen, dass die Mess-Items reduziert werden können. Da auch hier zusammenfassende Werte in die Item-Matrix eingebunden sind, ist es sinnvoll, im Hinblick auf die Validierungsstrategie die Einzelmesswerte zu verwenden und erneut zu überprüfen, inwiefern die Messwerte miteinander korrelieren. Genauer lassen sich diese hohen Anteile an bedeutsamen Korrelationen in den MMMT-Matrizen daher erklären, dass die im PSSUQ zu erhebenden Items zu vier Subskalen zusammengefasst werden, deren Werte ebenso in die Evaluierungsmatrizen eingeflossen sind. Demnach sind die errechneten Werte wie die Nützlichkeit des Systems, die Informationsqualität, die Qualität der Schnittstelle sowie der Wert zur Bestimmung der Gesamtgebrauchstauglichkeit aus dem Methoden-Inventar zu exkludieren.

Die Gegenüberstellung der konvergenten im Vergleich zu der diskriminanten Validität ist nur bezüglich der Abgrenzung zu dem Konstrukt Workload in der Machbarkeitsstudie nicht eindeutig. In der RD-Erhebung hingegen wird das Konstrukt Gebrauchstauglichkeit von den Konstrukten Workload und Situationsbewusstsein abgegrenzt. Im Unterschied zur RD-Evaluierung wurde der PSSUQ in der Machbarkeitsstudie zwei Mal mit je einem Design erhoben. Da diese Designs allerdings auch mit einer unterschiedlichen Bedienschwierigkeit einhergegangen sind, ist es naheliegend, dass dies die Abgrenzung zu Workload erschwert hat.

In den Erhebungen, in denen der PSSUQ verwendet wurde, zeigt sich dieser als geeignete Methodik der Gebrauchstauglichkeit, die zuverlässig von Situationsbewusstsein diskriminiert. Workload hingegen beeinflusst unter Umständen auch die Aussagen der Probanden, die mittels des PSSUQs erhoben werden.

8.2.3.3 *Eye-Tracking*

Zur Erfassung der Gebrauchstauglichkeit wurden neben der Augenaktivität als Indikator für die Suchstrategie des Nutzers auch zeitabhängige Blickmaße wie beispielsweise die Blickrate pro Sekunde oder die Gesamtblickdauer erfasst.

Unter Betrachtung der Eye-Tracking-Zeitleistung ist eine Kürzung des Methoden-Inventars angeraten aber nicht zwingend notwendig, da die Anteile an bedeutsamen Korrelationen in den MMMT-Matrizen eher moderat sind. Da sowohl Gesamtwerte sowie deren Mittelungen in das Methoden-Inventar eingeflossen sind, ist es allerdings sinnvoll, diese Werte im Hinblick auf die Validierungsstrategie zu entfernen.

Deutlich wird, dass die verwendeten zeitabhängigen Blickmaße die Methoden der Gebrauchstauglichkeit nicht eindeutig von den Methoden der anderen beiden Konstrukte abgrenzen können. In der vorliegenden Analyse wird offensichtlich, dass diese Blickmaße eher dazu geeignet gewesen wären,

Situationsbewusstsein zu erheben als Gebrauchstauglichkeit. Die Analyse der zeitabhängigen Eye-Tracking-Maße lässt daher auf eine enge Verbindung mit den Methoden des Situationsbewusstseins schließen, die sich insbesondere anhand der Nähe zu den Messungen wie beispielsweise der Fixationsdauer erklären lassen. Auch Workload wird in der Hälfte der Fälle besser beschrieben als die Methoden der Gebrauchstauglichkeit. Folglich sollten diese zeitabhängigen Eye-Tracking-Messungen nicht weiter im Rahmen der Gebrauchstauglichkeit erfasst werden. Messwerte, die jedoch besonders gut mit dem Situationsbewusstsein korrelieren, sind stattdessen zur Erhebung dieses Konstruktes besser geeignet und sollten daher hierzu genutzt werden.

Zur Evaluierung der Suchstrategie des Operators und der Informationsdispersion wurde die Augenaktivität des Nutzers im Rahmen des Konstruktes Gebrauchstauglichkeit sowohl global als auch aufgabenspezifisch erfasst. Nach Analyse beider Herangehensweisen im Hinblick auf die MMMT-Matrizen stellt sich heraus, dass eine Reduktion des Methoden-Inventars nur mit Vorsicht und unter genauerer Untersuchung der verwendeten Items vorzunehmen ist, da die globale sowie die aufgabenspezifische Herangehensweise sehr unterschiedliche Ergebnisse produzieren. Eine Reduktion des Methoden-Inventars auf die Verwendung der Messdaten von nur einem Auge ist jedoch plausibel, da in der RD-Studie, der 3DA-Evaluierung und der Sensor-Studie ein binokulares Eye-Tracking-System verwendet worden ist, dass die Mess-Items für das linke und das rechte Auge aufgezeichnet hat.

Die Gegenüberstellung der Validitäten führt des Weiteren zu der Schlussfolgerung, dass die Augenaktivität sowohl in der globalen als auch in der aufgabenspezifischen Herangehensweise nicht dazu geeignet war, die Konstrukte zu unterscheiden oder abzugrenzen. Es wird höchstens eine stärkere Bindung an das Konstrukt Workload in der globalen Erhebung der Augenaktivität sichtbar, die jedoch in der aufgabenspezifischen Herangehensweise nicht im gleichen Maße auffällt. Dass die Augenaktivität mit Situationsbewusstsein verbunden ist, ist anhand der messspezifischen Gegebenheiten zu erklären. Da eine Fixation das Auge in seiner Ruheposition und daher während der Informationsaufnahme erfasst, beschreiben Sakkaden die Augenbewegungen, die notwendig sind, um das Auge auf den nächsten visuellen Inhalt zu fokussieren (Bojko, 2013). Beide Messungen sind daher per Definition räumlich und zeitlich voneinander abhängig. Da sich die Mess-Items in den Untersuchungen allerdings unsystematisch verhalten, sollte generell von einer Weiterverwendung dieses Eye-Tracking-Parameters innerhalb des Methoden-Inventars abgesehen werden.

8.2.4 Zusammenfassende Betrachtung des Methoden-Inventars

Eine zusammenfassende Betrachtung des Methoden-Inventars offenbart, dass bei dessen Anwendung zum einen die Datenlage ökonomischer gestaltet werden kann, aber auch inhaltliche Aspekte beachtet werden müssen, die für die Weiterverwendung der Methoden und/oder deren Anpassung relevant sind.

Ein Großteil der verwendeten Methoden weist eine geringe Differenzierung der methodenimmanenten Items auf. Dies ist jedoch insbesondere auf die Verwendung von zusammenfassenden Werten wie sie beispielsweise beim NASA TLX, dem 3D SART sowie dem PSSUQ vorkommen und auf redundant erhobene Werte, wie sie unter Anwendung eines binokularen Eye-Tracking-Systems entste-

hen, zurückzuführen. Unter Weglassung dieser Datenpunkte könnte das Methoden-Inventar stringenter genutzt werden und so auch zu einer sensitiveren Anwendung hinsichtlich der Abgrenzung der Konstrukte führen.

Bei Betrachtung des gesamten Methoden-Inventars ist außerdem zu erkennen, dass die subjektiven Methoden wie der 3D SART für das Situationsbewusstsein, der NASA TLX zur Erhebung von Workload oder der PSSUQ zur Erfassung der Nutzerzufriedenheit für eine Weiterverwendung gut geeignet sind. Alle Fragebögen sind einfach zu verwenden, spiegeln die Befragungsergebnisse der Nutzer als Experten wider und erfassen die Konstrukte, in denen sie anzuwenden sind, ausreichend – wenn auch eher global. Insbesondere, wenn der Anwendungsaufwand der Informationsgewinnung gegenübergestellt wird, zeigt sich der Nutzen der subjektiven Fragebögen, die außerdem mit freien Kommentaren ergänzt werden können, um auch aufgabenspezifische Aspekte der Bedienung bzw. der Aufgabenbearbeitung durch den Nutzer erfassen zu können. Nachteilig ist, dass die globale Anwendung dieser Fragebögen zu einer Überschneidung mit den anderen Konstrukten führen kann – die Konstrukte also nicht eindeutig abgegrenzt werden können. Methodische Überschneidungen zwischen Workload und Situationsbewusstsein sind beispielsweise auch in der Verwendung des SART und des NASA TLX, wie sie auch Selcon et al. (1991) beschreiben, gegeben. Die enge Verbindung von PSSUQ und Workload ist in den vorliegenden Evaluierungen in dem Sinne zu erklären, dass die Aufgaben-Schwierigkeit, wie sie beispielsweise in der Bedienung von unterschiedlich gut geeigneten Designs entsteht, die Antwort-Tendenzen der Operateure im PSSUQ beeinflussen könnte. Hierzu sollten im Bereich der Cockpитеvaluierung weitere empirische Studien angedacht werden.

Die aufgabenspezifischen und funktionsrelevanten Messungen des Situationsbewusstseins wie der SAGAT, die Sortieraufgabe, die Gefahren-Detektierung und die aufgabenspezifischen Messungen des Eye-Trackings sind hinsichtlich ihrer Konstrukt-Differenzierung unterschiedlich gut geeignet, weil sie in ihrer Anwendungsform unterschiedlich gut für den eingesetzten Kontext genutzt werden konnten. So zeigt sich, dass die Anwendung des SAGATs in der RD-Evaluierung weniger gut geeignet ist als die Anwendung des SAGATs in der 3DA-Evaluierung. Ein direkter Vergleich der verwendeten Mess-Items ist aufgrund der Heterogenität der Items jedoch nicht möglich, da in der RD-Studie vielmehr die Erinnerungsleistung hinsichtlich der räumlichen Orientierung des Nutzers und dessen Urteilssicherheit im Fokus stand, während in die 3DA-Evaluierung neben der räumlichen Zuordnung und Wahrnehmung von Gefahren auch noch die Leistungsschätzungen des Nutzers sowie zeitliche Aspekte miteingeflossen sind. Insofern ist ein intrusiveres Vorgehen, das zeitlich gesehen näher am Untersuchungsgegenstand ist, jedoch auch eine geringere Anzahl an Items aufweist, in der Anwendung sinnvoller. Eine Weiterverwendung des SAGATs sollte in weiteren Versuchen überprüft werden. Allerdings sollte dies unter der Maßgabe geschehen, dass der Einsatz dieser Methode ökonomisch gerechtfertigt ist, da die aufgabenspezifische Anpassung dieser Methodik sehr zeitaufwendig ist. Weitere Methoden wie die Sortieraufgabe oder die Gefahren-Detektierung, die aufgabenspezifisch und als funktionsrelevante Maße entwickelt wurden und inhaltlich dem Konstrukt Situationsbewusstsein zuzuordnen sind, haben sich in den vorliegenden Untersuchungen als nicht geeignet erwiesen und sollten entweder ganz aus dem Methoden-Inventar entfernt oder einem anderen Konstrukt zuge-

ordnet werden. Letzteres ist insbesondere für die Sortieraufgabe möglich, da hier die Aufgabenstellung von den Probanden anscheinend weniger im Sinne der Handlungsplanung als vielmehr im Sinne der Handlungsdurchführung umgesetzt wurde und daher dem Konstrukt der Gebrauchstauglichkeit zuzuordnen ist. Bei Verwendung der Gefahren-Detektierung als Methodik ist jedoch die Anpassung an den Kontext für eine Weiterverwendung maßgeblich, zumal diese Methodik ähnliche Mess-Items erfragt, wie dies mittels SAGAT geschieht. Jedoch hat sich gezeigt, dass der Anteil an konvergenter Validität für diese Methodik in der 3DA-Evaluierung irrelevant war.

Hinsichtlich der Erfassung von Workload ist eine Weiterverwendung der TDRT in der vorliegenden Form nicht angezeigt, da eine genauere Bestimmung der Konstrukte sowie deren Abgrenzung nicht vorgenommen werden kann. Allerdings ist eine Neuauslegung der Methode im Sinne der Signalentdeckungstheorie unter Betrachtung der Sensitivität des Nutzers (s.a. McNicol, 2005) möglich, da so nicht nur ausgelassene Reize und die Reaktionszeiten des Nutzers erfasst werden können, sondern auch das Antwortverhalten der Versuchsperson sowie deren Diskriminierungsleistung als Sekundäraufgabe in Betracht gezogen werden kann. Eine Alternative hierzu ist es, die verwendete Methodik und die zugehörige Reizauslegung an die 2016 erschienene Norm ISO 17488:2016 anzupassen und zu überprüfen.

Eine exklusive Erfassung von Zeitleistung in Form von Bediengeschwindigkeiten und Reaktionszeiten als objektive Messmethodik für das Konstrukt Gebrauchstauglichkeit ist nicht eindeutig, da sich anhand dieser Methodik das Konstrukt nicht hinreichend bestimmen und abgrenzen lässt. Dies ist damit zu erklären, dass auch in den Konstrukten Situationsbewusstsein und Workload Reaktionszeiten in Form von Antwortzeiten und während der Bearbeitung von Zusatzaufgaben erfasst werden und diese Aufgaben beispielsweise auch anhand der Aufgaben-Schwierigkeit moderiert werden könnten. Ein weiteres Beispiel hierfür ist, dass ein generisches Leistungsmaß wie Reaktionszeit für die jeweiligen Konstrukte unterschiedlich verwendet werden kann. Die Bearbeitungszeit einer Aufgabe im Cockpit mit einem neuen Design kann einerseits dahingehend interpretiert werden, dass der Nutzer ein besonders gutes Situationsbewusstsein (1. Level Wahrnehmung) hat und deshalb besonders schnell reagieren konnte. Verändert man den Fokus jedoch, kann beispielsweise auch ein besonders gut integrierter und salienter Reiz dazu geführt haben, dass ein Nutzer schneller reagiert oder eine kürzere Bedienzeit aufweist. Daher ist es sinnvoll, die Bedienzeit für die Validierungsstrategie – nicht aber hinsichtlich der Erhebung der Konstrukte aus dem Methoden-Inventar zu exkludieren.

Die Verwendung von Eye-Tracking als objektive Methodik erweist sich insofern als kritisch, als dass es in allen drei Konstrukten verwendet wird. Hierbei lassen sich vor allem die Workload-Maße abgrenzen, da diese insbesondere die Pupille oder den Lidschluss erfassen. Beide Maße weisen jedoch sehr niedrige Anteile an konvergenter Validität auf und sollten daher nicht weiter im Methoden-Inventar verwendet werden. Dies ist bei der Erfassung des PERCLOS durchaus wünschenswert, da dieses Maß zur Erhebung von Ermüdung genutzt wird (s.a. Wierwille et al., 1994) und nicht Workload erfassen sollte. Die Variable Pupillen-Dilatation zeigt jedoch, dass dieses Maß nicht geeignet ist, um Workload zu erfassen. Dies kann darin begründet sein, dass die Simulationsumgebung als solches nicht dazu geeignet war, um die durch Workload induzierte Erweiterung der Pupille zu erheben, da

die Projektion des Szenarios Beleuchtungsschwankungen verursachte. Entsprechend weisen beispielsweise auch Bartels & Marshall, 2012, darauf hin, dass dieses Maß bei Beleuchtungsveränderungen störanfällig ist. Hinsichtlich der verbleibenden Konstrukte wie Situationsbewusstsein und Gebrauchstauglichkeit kommt hinzu, dass hier sequentiell abhängige Maße wie Blicke und Transitionen sowie Fixationen und Sakkaden zur Erhebung von jeweils unterschiedlichen Konstrukten verwendet worden sind, die im Rahmen der Validierungsstrategie Überschneidungen der Messvorgehen aufzeigen. Hinzu kommt, dass Eye-Tracking in der Anwendung in der Simulationsumgebung wenig effizient einsetzbar ist, da die Integration in die Simulationsumgebung langwierig ist, die Aufzeichnung der Daten störanfällig ist und das Datenmaterial selbst zeitaufwendig aufbereitet werden muss, bevor eine Ableitung der Ergebnisse möglich ist. Eine Verwendung des Eye-Trackings sollte daher nur dann angezeigt sein, wenn dieses auch zielführend eingesetzt werden kann, also die Erhebung des Blickes dazu führt, dass ein bestimmtes Verhalten (wie beispielsweise die Beobachtung eines Gefahrenreizes in der 3DA-Evaluierung) so gemessen werden kann und kein einfacheres oder gängigeres Mittel verfügbar ist (s.a. Bojko, 2013).

Insgesamt gesehen wird also offensichtlich, dass für die Erhebung der Auswirkungen von Designänderungen beispielsweise in einem Cockpit-Simulator weiter ein sinnvoller Methodenmix notwendig ist, um die Vorteile der Einzelmethoden umfänglich zu nutzen und deren Nachteile auszugleichen. Eine Verwendung von subjektiven und objektiven Methoden ist notwendig, da nur sie die Expertenmeinung wiedergeben und in Kombination ein vollständiges Bild für weitere Design-Entscheidungen zeichnen. Hierbei sollte Eye-Tracking nur sparsam und zielführend als objektive Methode verwendet werden, da die erfassten Maße zum Teil doppeldeutig und in verschiedenen Konstrukten gemessen werden, ohne diese zu differenzieren. Weiterhin sollten in das Methoden-Inventar Methoden integriert sein, die sowohl globale Aspekte des Konstrukts erfassen als auch Methoden, die aufgabenspezifische Aspekte erheben, um die jeweiligen Konstrukte vollständig zu beschreiben. Dies ist insbesondere für das Konstrukt Situationsbewusstsein von großer Bedeutung, da beispielsweise eine globale Methode wie der SART nicht mit einer aufgabenspezifischen Methode wie dem SAGAT korreliert (Endsley et al., 1998; Salmon et al., 2007), da beide Methoden unterschiedliche Aspekte des Situationsbewusstseins beschreiben (Salmon et al., 2007).

Ergänzend muss betont werden, dass für die Vollständigkeit eines erfolgreich angewendeten Methoden-Inventars die qualitativen Methoden wie Interviews oder beispielsweise der Cognitive Walkthrough von großer Bedeutung und unter Umständen sogar unerlässlich sind. Sie sind deshalb nahezu unverzichtbar, da diese Herangehensweisen eine Informationsvielfalt erfassen, die zum einen die Bewertung des Nutzers erklären, den Nutzungskontext besser würdigen und Designfehler spezifizieren sowie mögliche Designvorschläge systematisch generieren. Auf Grund ihrer qualitativen Beschaffenheit konnten diese Methoden jedoch nicht in die gewählte Validierungsstrategie integriert werden.

9 ANALYSE UND BEWERTUNG DES MULTI-TRAIT/MULTI-METHOD-ANSATZES ALS VALIDIERUNGSSTRATEGIE

Neben den methodischen Implikationen und deren Auswirkungen auf die Qualitätskriterien, die in Kapitel 8 beschrieben werden, ist auch der Multi-Trait/ Multi-Method-Ansatz als Validierungsstrategie hinsichtlich seiner Eignung zu diskutieren. Die Einflussfaktoren, die hierzu betrachtet werden können, ergeben sich aus der Definition der Qualitätskriterien und deren theoretischer Abgrenzbarkeit, der Beschaffenheit des Methoden-Inventars und der hierin zu Grunde liegenden Datenstruktur sowie der Beschaffenheit der Validierungsstrategie an sich.

Bei Überprüfung der Konstruktvalidität wird der Fokus auf die Eignung von Messinstrumenten gelegt (Diekmann, 2005), um theoretische Zusammenhänge eines Konstruktes mit anderen zu erklären und diese anhand der verwendeten Messinstrumente empirisch zu prüfen (Schnell, Hill & Esser, 2018). Für die Konstrukt-Validierung des Methoden-Inventars wird somit offensichtlich, dass die Qualitätskriterien als theoretische Konstrukte zu verstehen sind, die sich durch ein Netzwerk an Assoziationen und deren Propositionen zueinander zusammensetzen (Cronbach & Meehl, 1955). Dieses Netzwerk an Assoziationen schlägt sich in der eingesetzten Methodenvielfalt für jedes Konstrukt und der zugehörigen Datenstruktur nieder. Hinsichtlich der Validierungsstrategie stellt sich somit die Frage, ob die Ergebnisse aus Kapitel 7 und der folgenden, in Kapitel 8 geführten Konstrukt- und Methodendiskussion dazu geeignet ist, um zu diesen Schlussfolgerungen zu gelangen und inwiefern die diskutierten Ergebnisse von der Beschaffenheit der Validierungsstrategie selbst beeinflusst wurden.

Unter Betrachtung der Ergebnisse aus Kapitel 7 zeichnet sich ab, dass die Anteile an bedeutsamen Korrelationen sowohl hinsichtlich der konvergenten wie auch der diskriminanten Validität niedriger werden, je mehr Methoden sukzessive zur Beschreibung eines Konstruktes eingesetzt werden. Dies wird ebenso deutlich, wenn die in der Evaluierung verwendeten Datensätze umfangreicher sind wie beispielsweise aufgrund eines größeren Stichprobenumfangs. Insbesondere ersteres ist daher auch ein Grund, warum in der Machbarkeitsstudie Situationsbewusstseins besonders gut beschrieben zu sein scheint. Da hier Situationsbewusstsein anhand einer globalen und einer aufgabenspezifischen Messung mittels Eye-Tracking bestimmt wurde, ist so der besonders hohe Anteil an konvergenter Validität zu erklären, der in allen anderen vorhandenen Evaluierungen so nicht vorliegt. Folglich sollte die konvergente Validität mit mindestens zwei verschiedenen Methoden und unterschiedlichen Messinstrumenten erhoben werden, um die konvergente Validität des untersuchten Konstruktes und die diskriminante Validität bei Untersuchung anderer Konstrukte nicht zu überschätzen. Eine weitere Schlussfolgerung hieraus ist es, hinsichtlich der Validierungsstrategie auf eine Ausgeglichenheit der verwendeten Methoden zu achten. Insbesondere die Nutzung des Eye-Trackings mit einer Vielzahl an unterschiedlichen Mess-Items in allen drei Konstrukten verzerrt hier die Ergebnisse der Evaluationsmatrizen aufgrund der messimmanenten Zusammenhänge. Eine ausgewogene Evaluationsmatrix sollte daher verschiedene subjektive Methoden, objektive Methoden sowie aufgabenspezifische und globale Methoden für jedes Konstrukt aufweisen. Allerdings zeigt sich gerade für die aufgabenspezifischen Methoden wie beispielsweise dem SAGAT die Problematik, dass diese nur schwer studien-

übergreifend verglichen und auf ihre Eignung hin untersucht werden können, weil aufgrund der Aufgabenspezifität teils unterschiedliche Aspekte des Konstruktes gemessen werden könnten. Auch deshalb sollten in das zu untersuchende Methodenkonvolut möglichst wenig Methoden einfließen, die aufgabenspezifisch und aus dem Nutzungskontext heraus neu entwickelt wurden (wie beispielsweise Sortieraufgabe, Gefahren-Detektierung). Diese Methoden können nur schwer studienübergreifend verglichen werden und verzerren möglicherweise den Anteil an konvergenter Validität bei den etablierten Methoden. Auch der Einsatz des PERCLOS als Methodik zur Erfassung von Ermüdung verzerrt insofern die Evaluationsmatrizen als er die verbleibenden Methoden zur Erhebung von Workload hinsichtlich deren konvergenter Validität schwächt. Allerdings hat sich gerade im Hinblick auf diese Methodik der Einsatz des Multi-Trait/ Multi-Method-Ansatzes als Validierungsstrategie als besonders geeignet erwiesen, da mit Hilfe der Validierungsstrategie die mangelnde Zugehörigkeit dieser Methodik zu den Konstrukten kontrastiert wird.

Neben der Methoden-Kombination und der ihr zu Grunde liegenden Datenstruktur hat auch die Beschaffenheit der Validierungsstrategie Auswirkungen auf deren Aussagekraft. Ein wesentlicher Aspekt ist, dass die Daten, die mit dem Methoden-Inventar gemessen worden sind, in eine Korrelationsmatrix überführt und gemäß den Konstrukten und ihrer Methoden zusammengefasst werden. Dies führt zu einem Informationsverlust auf zwei Ebenen – der Daten-Ebene und der Methoden-Ebene. Die Daten-Ebene ist betroffen, da die Evaluationsmatrizen aus Rangkorrelationen nach Spearman bestehen, weil in das Methoden-Inventar neben intervallskalierten Messungen auch ordinale Methoden eingegangen sind. Dies fällt allerdings nicht so stark ins Gewicht, da ein oberflächlicher Vergleich der Korrelationskoeffizienten von Spearman und Pearson zu ähnlichen Ergebnissen führt. Da die Daten zudem als Anteile an bedeutsamen Korrelationen ausgewertet und die jeweiligen Items zu Methoden verbunden worden sind, sind außerdem die Korrelationen als Betrag erfasst worden. Das heißt, dass jegliche Aussage, in welche Richtung verschiedene Mess-Items oder deren Methoden zusammenhängen, nicht getroffen werden kann, sofern die Matrix als Ganzes ausgewertet wird.

Weiterhin müssen die Daten zu einer Methodik zusammengeführt und an die erforderliche Datenstruktur angepasst, das heißt reduziert, werden. Dies betrifft insbesondere das Eye-Tracking als methodische Herangehensweise. Die Eye-Tracking-Daten wurden während den Evaluierungen nicht nur global und aufgabenspezifisch erfasst, sondern auch inhaltsabhängig, mit Bezug auf die für die Mensch-Maschine-Interaktion (MMI) relevanten, jedoch verschiedenen Displayinhalte. Diese Daten sind für jede Evaluierungsmatrix separat zur Bildung der Korrelationen zu einem Wert zusammengefasst worden. Dies führt dazu, dass evtl. durch die Displayinhalte mediierte Unterschiede im Blickverhalten ignoriert und somit einen weniger bedeutungsvollen Wert bilden als beispielsweise die Werte des NASA TLX oder anderer Methoden in ein und derselben Versuchsbedingung. Für die Validierungsstrategie heißt das, dass aufgrund der erforderlichen Datenstruktur Werte mit unterschiedlicher Qualität in die Evaluierungsmatrix eingegangen sind und hier zu einer Unterschätzung des Eye-Trackings beigetragen haben könnten.

Die durch den Multi-Trait/ Multi-Method-Ansatz geforderte Datenstruktur hat des Weiteren den Nachteil, dass nur quantifizierbare Methoden aufgenommen und validiert werden können. Qualitati-

ve Methoden hingegen wie der Cognitive Walkthrough oder die Erkenntnisse, die anhand der vorgenommenen Interviews erzielt wurden, können hier nicht validiert werden, obwohl diese Methoden während der Evaluierung wesentlich zur Bestimmung der Nutzerzufriedenheit und der Offenlegung von Designfehlern und Nutzerstrategien im Anwendungskontext beigetragen haben und – nicht zuletzt – für die abschließende Interpretation der MMI-Befunde bedeutsam sind.

Auch die Analyse und Interpretation der Korrelationskoeffizienten der Evaluationsmatrizen im Sinne des Multi-Trait/ Multi-Method-Ansatzes bergen Kritikpunkte, auf die Schermelleh-Engel und Schweizer (2012) hinweisen:

- 1) Die Korrelationskoeffizienten werden im Sinne von Häufigkeiten ausgewertet
- 2) Mit manifesten Variablen werden latente Traits und Methoden beschrieben
- 3) Konvergente und diskriminante Validität sind auf Korrelationsebene miteinander konfundiert
- 4) Die Auswertung der Korrelationskoeffizienten ist deutungsabhängig, d.h. subjektiv

Unter 1) verweisen die Autoren darauf, dass Korrelationskoeffizienten üblicherweise unter Angabe der zugehörigen Konfidenzintervalle als größer oder kleiner interpretiert werden und nicht auf Grund einer Häufigkeitszählung. Dies liegt darin begründet, dass Konfidenzintervalle mit dieser Strategie nicht interpretiert werden können, weil diese nicht wie die Korrelationskoeffizienten mittels Häufigkeit zusammengefasst werden können. Auch die zugehörigen Signifikanzen können nicht für weiter Interpretationen in Betracht gezogen werden, da diese auch nicht bezüglich der jeweiligen Methoden zusammengefasst werden können, weil sie sich nur auf spezifische Item-Paarungen beziehen. Der zweite Kritikpunkt 2) lässt sich insofern relativieren, als dass das vorliegende Vorgehen hauptsächlich dazu genutzt worden ist, explorativ die Eignung von Human-Factors-Methoden für die Untersuchungen im Cockpit-Simulator zu überprüfen. Auch Campell und Fiske (1959) interpretieren ihr Vorgehen als unterstützende Strategie, um die Eignung der Methoden zur Beschreibung der Konstrukte zu bestimmen. Rückschlüsse auf latente Aspekte der Konstrukte werden daher in Anbetracht der verwendeten Methoden vorgenommen und zeigen lediglich Überschneidungen der Konstrukte, ergeben jedoch keine konkreten Hinweise zur Reformulierung der Konstrukte unter Analyse des gezeigten Verhaltens der Probanden. Allerdings können die vorliegenden Ergebnisse dazu genutzt werden, um diese Hypothesen zu bilden und in weiteren Experimenten (siehe Kapitel 10.1) strukturiert zu untersuchen. Der dritte Kritikpunkt 3) zielt ebenso auf die Verwendung des Korrelationskoeffizienten zu Analyse Zwecken ab. Bezüglich des Multi-Trait/ Multi-Method-Ansatzes bezieht sich dies darauf, dass der Methodeneinfluss auf die Traits anhand der Korrelation in der Matrix inhärent ist und dies über die Bestimmung von konvergenter und diskriminanter Validität konfundiert ist. Das heißt, dass die zentrale Idee von Campbell und Fiske (1959), nämlich die Beschreibung der Auswirkung der verwendeten Methoden im Sinne von konvergenter und diskriminanter Validität, nach Schermelleh-Engel und Schweizer (2012) zwar an sich als geeignet angesehen wird, die Korrelationsmatrix jedoch kritische Aspekte hinsichtlich der Auswertung aufweist. Dies zeigt sich auch unter Punkt 4), der als kritisch aufzeigt, dass die Analyse der Korrelationsmatrix subjektiv ist, da von Campbell und Fiske kein festes Vorgehen zur Analyse der Matrizen vorgegeben wird. Hinsichtlich der Analyse der Evaluationsmatrizen wurden die relativen Häufigkeiten der MonoMethod-MonoTrait-, den Hetero-

Method-MonoTrait- und den HeteroMethod-HeteroTrait-Matrizen einheitlich nach bedeutsamen, weniger bedeutenden und irrelevanten Korrelation zusammengefasst. Auch die Eignung der Methoden hinsichtlich der konvergenten und diskriminanten Validität wurde dahingehend beurteilt, in welchem Verhältnis sich diese zueinander verhalten, je nachdem, welche Validitätsform höhere Anteile aufgezeigt. Gerade in Bezug auf den letzten Aspekt ist daher ein subjektiver Einfluss zur Auswertung vorhanden. In Bezug auf die Verwendung des Korrelationskoeffizienten hingegen ist ein einheitliches Entscheidungskriterium gewählt worden, welches explorativ und systematisch die Mess-Items und Methoden aufzeigt, die sich für eine Anwendung im Cockpit-Simulator eignen. Die Verwendung von alternativen Methoden zur Analyse von Multi-Trait/ Multi-Method-Matrizen wie beispielsweise mittels konfirmatorischen Faktorenanalyse anhand von Correlated-Trait-Correlated-Method-Modellen oder Correlated-Trait-Correlated-Uniqueness-Modellen (Bühner, 2011) war hingegen nicht angezeigt, da die hierfür notwendigen Voraussetzungen nicht erfüllt waren. Im vorliegenden Kontext wurden die Daten mit Hilfe einer Experten-Stichprobe erhoben, weil deren Verhalten im Nutzungskontext gemessen werden sollte und die Grundgesamtheit dieser Experten klein ist (Rabl, 2016). Die Forderungen der konfirmatorischen Faktorenanalyse nach einer großen Stichprobenzahl zur Erhebung einer entsprechenden Datenbasis konnte aus diesem Grund nicht erfüllt werden, weil selbst eine Vollerhebung in dieser eingeschränkten Grundgesamtheit nicht zu genügend vielen Probanden führen würde.

Die Kritik an dem Multi-Trait/ Multi-Method-Ansatz ist, wie die geführte Diskussion zeigt, in Teilen durchaus gerechtfertigt. In der Zusammenschau zeigt sich somit, wie auch Messick bereits (1987) beschreibt, dass die Validierungsstrategie anfällig hinsichtlich der Daten-Erhebung, der Reliabilität der zur Verfügung stehenden Messinstrumente und ihrer Messungen sowie der Abgrenzbarkeit und Bedeutungsreichweite der Konstrukte ist. Der Autor sieht den Multi-Trait/ Multi-Method-Ansatz daher als heuristisches Vorgehen, das den Untersuchenden dazu zwingt, konvergente und diskriminante Validität oder deren Fehlen gleichzeitig zu adressieren, zu betrachten (Messick, 1987) und damit eine Gesamtbewertung vorzunehmen.

Die Verwendung des Multi-Trait/ Multi-Method-Ansatzes zur Validierung der Methoden des Methoden-Inventars in einer Cockpit-Simulationsumgebung erweist sich grundsätzlich als geeignet, zeigt jedoch auch einige kritische Aspekte auf.

So können anhand dieses heuristischen Validierungsvorgehens die gewonnenen methodischen Erkenntnisse aus den Untersuchungen der Evaluationsmatrizen hinsichtlich ihrer Weiterverwendung oder eventueller Abwandlung für die Cockpit-Simulationsumgebung sowie als Grundlage für weitere Untersuchungen adäquat genutzt werden. Als Validierungsgrundlage wurden vier Korrelationsmatrizen genutzt und es zeigt sich über die vier Evaluierungen hinweg, dass die Abgrenzbarkeit der Konstrukte über die verwendeten Methoden beschrieben werden kann und des Weiteren die Methoden über die einzelnen Items. Es stellte sich somit in Kapitel 7 und 8 heraus, dass die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit als Konstrukte nicht ausreichend anhand der verwendeten Methoden differenziert werden konnten, da die eingesetzten Methoden eine

zu geringe konvergente Validität, dafür aber eine zu hohe diskriminante Validität aufweisen. Dies konnte anhand der Verwendung der Validierungsstrategie systematisch aufgezeigt werden. Daher wird offensichtlich, dass der Multi-Trait/ Multi-Method-Ansatz ein geeignetes Instrument zur Validierung von einem Methoden-Inventar ist, das mit Experten in einer Simulationsumgebung evaluiert wurde.

Zukünftig gibt es jedoch bei der Verwendung des Multi-Trait/ Multi-Method-Ansatzes auf Grundlage einer Experten-Stichprobe in einer Cockpit-Simulationsumgebung einige Aspekte, die beachtet werden sollten. So ist zum einen zu berücksichtigen, dass eine methodische Ausgeglichenheit der Matrix vorliegt; so sollten Messungen wie beispielsweise das Eye-Tracking im Vergleich zu den anderen Methoden nicht übermäßig, sondern in ähnlichem Umfang eingesetzt werden. Ebenso ist ein Augenmerk auf ein adäquates Verhältnis von aufgabenspezifischen Methoden zu globalen Methoden innerhalb der Erhebung der Konstrukte zu legen. Zum anderen sollten für die Validierungsstrategie zusammenfassenden Werte vermieden und universell einsetzbare Messungen sparsam eingesetzt werden. Werden diese Themen bei der Aufstellung der Korrelationsmatrizen berücksichtigt, ist zu erwarten, dass sich der Multi-Trait/ Multi-Method-Ansatz auch in weiteren Evaluierungen als systematische Validierungsstrategie erweisen wird.

10 OFFENE FORSCHUNGSFRAGEN

Im Hinblick auf die vorliegenden Evaluierungen und die verwendete Validierungsstrategie ergeben sich zwei Ansatzpunkte für weitere Forschungen. Zum einen erfordern die in Kapitel 8 vorgeschlagenen Methoden-Änderungen und -Anpassungen eine erneute experimentelle Überprüfung (Kapitel 10.1) und zum anderen können auch die zu kritisierenden Aspekte der Validierungsstrategie bzw. der Evaluationsmatrizen aus Kapitel 9 modifiziert und erneut bewertet werden (Kapitel 10.2).

10.1 Methodische Anpassungen

In der in Kapitel 8 geführten Diskussion hinsichtlich der Eignung der verschiedenen Methoden zur Beschreibung der Qualitätskriterien wurden einige methodische Aspekte sichtbar, die eine tiefere Untersuchung als weiteres Vorgehen erfordern.

Bezüglich des Situationsbewusstseins stellt sich hier die Gegenüberstellung von globalen Methoden und aufgabenspezifischen Methoden zur Schaffung einer ausgeglichenen Matrix als Ansatzpunkt für weitere Forschung dar. Hierzu sollten verschiedene validierte globale Methoden wie der SART, die ein festgelegtes Fragenset aufweisen wie beispielsweise die Crew Awareness Rating Scale (s.a. McGuiness & Ebbage, 2002), den aufgabenspezifischen Methoden wie dem SAGAT, auch Methoden der Echtzeitbefragung wie dem Situation Present Assessment Method (Durso et al., 1998) gegenübergestellt werden. Herauszufinden ist, inwiefern diese methodischen Konzepte eigentlich das theoriegeleitete Konstrukt Situationsbewusstsein abbilden. Allerdings besteht hinsichtlich dieser Untersuchung die Problematik, dass sich gerade die aufgabenspezifischen Methoden aufgrund des Nutzungskontextes inhaltlich sehr ähnlich sein werden, was zur Folge hat, dass nur ein Between-Subject-Design für die Evaluierung möglich ist und hierzu eine entsprechende Stichprobengröße erforderlich ist. Wird aber eine Untersuchung in diesem Rahmen durchgeführt, wäre es zudem möglich, für die weitere Interpretation auf die von Salmon et al. durchgeführte Inhaltsanalyse zurückzugreifen. Allerdings müssten auch hier die aufgabenspezifischen Aspekte verallgemeinernd auf eine abstraktere Ebene hin transformiert werden, um diese theoriegeleitet zu analysieren, um spezifische Schlüsse hinsichtlich des Konstrukts Situationsbewusstsein ziehen zu können. Bezüglich der vorliegenden Validierungsstrategie wäre es allerdings schon ausreichend, ein ausgeglichenes Verhältnis von globalen zu aufgabenspezifischen Methoden zu schaffen und deren Zusammenhänge auf Item-Ebene unter Verwendung eines geeigneten Nutzungskontexts zu überprüfen.

Die TDRT wurde als eher ‚junge‘ Methode in das Methoden-Inventar zur Erhebung von Workload integriert und ist daher nicht voll umfänglich an der ISO-Norm 17488 (2016) ausgerichtet, da dieser in der Evaluierung beispielsweise als dreimaliger taktiler Reiz appliziert wurde, bevor die Probandenreaktionszeit gemessen wurde. Für eine weitere Untersuchung erscheint es daher lohnenswert, den Reiz normgerecht, d.h. mit einer maximalen Reizdauer von einer Sekunde in einem zufälligen Intervall zwischen drei und fünf Sekunden (ISO 17488:2016), einzusetzen und diese Ergebnisse hinsichtlich der Erfassung von Workload in einer Cockpit-Simulationsumgebung erneut auf Eignung außerhalb

des Automobilkontextes zu überprüfen. Ein anderer Aspekt zur Weiterentwicklung der TDRT ist die mit der Signalentdeckungstheorie (s.a. Goldstein, 1999) kompatible Modellierung, da dieses Vorgehen auch das Probandenverhalten genauer kennzeichnet, wenn dieser einen Reiz zurückweisen muss. In der vorliegenden Version ist unklar, ob das Auslassen eines Reizes auf die vom Probanden empfundene Workload zurückzuführen ist oder ob eine bewusste Entscheidung des Probanden vorlag, auf diesen Reiz nicht zu reagieren. Das heißt, dass eine Umsetzung dieser Methodik in Form der Signalentdeckungstheorie zum einen eindeutiger zu interpretieren ist und zum anderen der Einfluss der Workload auf die Sensitivität der Probanden gemessen werden kann.

Ein weiteres Maß, das im Rahmen des Methoden-Inventars zur Beschreibung der Workload genutzt wurde, ist die Pupillen-Dilatation, deren Weiterverwendung als Workload-Maß im Hinblick auf die in Kapitel 8.2.2.3 geführte Diskussion nicht empfohlen werden kann. Ein entscheidender Nachteil dieses Parameters ist dessen Empfindlichkeit bezüglich schwankender Lichtverhältnisse (s.a. Bartels & Marshall, 2012; Bojko, 2013; Schwalm, 2009), die in einer Simulationsumgebung nur schwerlich zu kontrollieren sind. Um zu klären, inwiefern die vorhandenen Ergebnisse der Pupillen-Dilatation durch die Lichtverhältnisse in der Simulationsumgebung verfälscht sind, wäre ein vergleichendes, hypothesengeleitetes Experiment in der Simulationsumgebung unter Betrachtung der Pupillen-Dilatation und dem Index of Cognitive Activity (ICA) sowie einer weiteren validierten Workload-Methode wie beispielsweise dem NASA TLX sinnvoll. Der ICA, nach einer Patentschrift von Marshall (1999) ein Eye-Tracking-Maß, das aufgrund einer Wavelet Analyse diejenigen Pupillenerweiterungen herausfiltert, die durch besonders kurzfristige und schnelle Erweiterungen der Pupille zu Stande kommen. Zudem ist der ICA gegenüber sich verändernden Beleuchtungseinflüssen unempfindlich (Schwalm, 2009) und korreliert mit einer sich verändernden Aufgaben-Schwierigkeit (Marshall, Pleydell-Pearce & Dickson, 2003). Schwalm fand (2009) zudem heraus, dass sich der ICA auch in einem anwendungsnahen Kontext, wie einer simulierten Spurwechselfaufgabe, bewährt und Veränderungen der Workload beschreibt. In diesem Zusammenhang stellt sich die Frage, ob sich der ICA als Methode zur Erhebung von Workload auch in der vorliegenden Simulationsumgebung eignet und inwiefern die vorhandenen Ergebnisse auf die Beleuchtungseinflüsse des Simulators zurückzuführen sind. Ein vergleichendes Experiment beider Eye-Tracking-Maße zur Erfassung der Pupillen-Dilatation könnte hierzu Aufschluss geben.

Auch das Konstrukt Gebrauchstauglichkeit kann mit weiteren Methoden wie beispielsweise der System Usability Scale (Brooke, 1996; s.a. Stanton et al., 2011) ergänzt werden, um dahingehend zu solideren Validierungsergebnissen zu gelangen. Aus ökonomischen Gründen war dies allerdings im vorliegenden Vorgehen nicht möglich, da sich hier die inhaltliche Befragung im Nutzungskontext zu sehr mit dem PSSUQ überschneiden hätte. Zudem bleibt weiterhin der Nachteil bestehen, dass qualitative Methoden wie der oben beschriebenen Cognitive Walkthrough nur teilweise oder nicht in die Matrix integriert werden können. Außerdem gilt es, hinsichtlich des Qualitätskriteriums Gebrauchstauglichkeit, die offenbar gewordene Überschneidung von PSSUQ und Workload zu überprüfen. So haben die oben beschriebenen Ergebnisse die Vermutung nahegelegt, dass der PSSUQ auch durch die Aufgaben-Schwierigkeit im Nutzungskontext unter Bedienung der Cockpit-Funktion mode-

riert werden könnte. Dies müsste jedoch in der vorliegenden Versuchsumgebung überprüft werden, da sowohl Kokini, Lee, Richard und Moon (2012) hinsichtlich ihrer vorgenommenen Manipulationen von Workload und Nutzerkontrolle als auch Sonderegger und Sauer (2009) mit der Manipulation des Experimentalsettings unter Einführung von Beobachtern dem PSSUQ eine gewisse Robustheit nachweisen konnten.

Ein ergänzendes Vorgehen auf Grundlage der vorliegenden Ergebnisse wäre es, die Item-Ebene der Evaluationsmatrizen nochmals separat und losgelöst von den verwendeten Methoden zu analysieren. Die Items, die den größten Anteil an bedeutenden Korrelationen für ihr jeweiliges Konstrukt aufweisen, könnten nochmals theoriegeleitet überprüft und möglicherweise zu einer neuen Methode zusammengefügt werden – deren Eignung dann erneut in einem breiteren Nutzungskontext empirisch getestet werden müsste. Für dieses Vorgehen wäre es jedoch sinnvoll, die in Kapitel 9 genannten Kritikpunkte zur Umorganisation der Matrix zu berücksichtigen.

Abschließend ist es außerdem möglich, sofern ein stabiles Methoden-Set zur Beschreibung der Qualitätskriterien gefunden wird, dieses sukzessive mit weiteren Methoden zu ergänzen und deren Validität explorativ zu überprüfen. Die sich hieraus abzuleitenden Hypothesen können dann wiederum experimentell untersucht werden und somit eindeutiger auf das jeweilige Konstrukt bezogen werden.

10.2 Anpassungen der Validierungsstrategie

Ein einfaches Vorgehen zur weiterführenden Optimierung der Validierungsstrategie ist, wie oben bereits angedeutet, die Umstrukturierung der Evaluierungsmatrizen unter Berücksichtigung der unter Kapitel 9 aufgeführten Kritikpunkte in der Datenstruktur. Demnach sollten die zusammenfassenden Werte aus der Matrix entfernt werden, um übermäßige Anteile an bedeutsamen Korrelationen aus den MonoMethod-MonoTrait (MMMT)-Matrizen zu umgehen. Weiterhin müssten auch möglicherweise redundante Werte wie Eye-Tracking-Werte des linken und des rechten Auges aus der Evaluationsmatrix exkludiert werden. Dies würde auf Seiten der MMT-Matrizen dafür sorgen, dass die verwendeten Mess-Items innerhalb der Methoden differenzierter sind und die Konstrukte selbst somit eindeutiger beschreiben – womöglich mit der Folge, dass die Anteile an bedeutsamen Korrelationen insbesondere hinsichtlich der konvergenten Validität, aber auch der diskriminanten Validität weiter sinken. Auch könnte eine Optimierung der Evaluierungsmatrizen bedeuten, dass die Eye-Tracking-Items spezifischer und selektiver eingesetzt werden müssen, um ein Übermaß an Eye-Tracking in den Evaluierungsmatrizen zu verhindern und die Querabhängigkeiten dieses Mess-Instrumentes zu umgehen. Demnach sollten hinsichtlich Fixationen und Sakkaden entweder nur eines dieser beiden Maße verwendet werden oder beide Maße für ein einziges Konstrukt genutzt werden. Letzteres gilt insbesondere für die Konstrukte Gebrauchstauglichkeit und Situationsbewusstsein. Die Betrachtung von Workload ist hiervon etwas losgelöst, da für die Erfassung dieser Parameter andere inhalts- und zeitunabhängige Maße wie beispielsweise die Pupillenerweiterung erhoben worden sind. Ergänzend müsste für dieses Qualitätskriterium der PERCLOS entfernt werden, da dieser Messwert Ermüdung erhebt und somit für die Erfassung von Workload ungeeignet ist und hierdurch

die Evaluationsmatrix verzerrt. Auf Grundlage der Umorganisation der Evaluierungsmatrizen wären die Aussagen der Validierungsstrategie und deren Eignung iterativ weiter überprüfbar.

Zielführend erscheint auch eine Gegenüberstellung der korrelativen Evaluierungsmatrizen und der Verwendung einer konfirmatorischen Faktorenanalyse unter der Voraussetzung, dass eine ausreichende Datenbasis vorhanden ist. Nach Schermelleh-Engel und Schweizer (2012) ist die konfirmatorische Faktorenanalyse am besten dazu geeignet, die Grundideen von Campell und Fiske fortzuführen, da hier Trait und Methodenauswirkungen separat untersucht werden können. Bühner (2011) verweist hier auf zwei unterschiedliche Formen der konfirmatorischen Faktorenanalyse – die Correlated-Trait-Correlated-Method-Modelle und die Correlated-Trait-Correlated-Uniqueness Modelle. Allerdings erfordern beide Validierungsstrategien eine ausreichende Datenbasis und insbesondere die Correlated-Trait-Correlated-Method Modelle könne zu Schätzproblemen führen, da hier Modellparameter teils mit Konstanten gleichgesetzt werden müssen. Bei den Correlated-Trait-Correlated-Uniqueness Modellen hingegen wird eine Methode als Standardmethode ausgewählt und die Methodenfaktoren nur für die Nicht-Standardmethoden modelliert (Bühner, 2011). Für die vorliegende Arbeit und für die Evaluation im Cockpit-Simulator mit einer Experten-Stichprobe ist jedoch dieser Vergleich zur konfirmatorischen Faktorenanalyse auf Grund der zur Verfügung stehenden Stichprobengröße nicht herzustellen; selbst eine Vollerfassung der Experten könnte nicht die erforderliche Anzahl von Datenpunkten liefern.

11 ZUSAMMENFASSUNG

Bei der Implementierung neuer Cockpit-Funktionen muss die Qualität der Mensch-Maschine-Schnittstelle hinsichtlich der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit auch bei kleinem Nutzerkreis neu bewertet werden. Dies kann nur unter bestimmten Voraussetzungen geschehen.

Da Cockpits von hochagilen Flugzeugen eher einen kleinen Nutzerkreis haben, muss die Qualität der Schnittstelle innerhalb der Stichprobe inhaltlich umfänglich und mit verschiedenen Messwerkzeugen – also einem Methodenmix – bestimmt werden. Eine weitere Voraussetzung ist es, die Komplexität der Aufgaben im Cockpit so zu reduzieren, dass die Funktionalität in einer repräsentativen Flugaufgabe erfasst wird. Das heißt, dass diese spezifische Messsituation kontrollierbar und somit die Funktionsnutzung selbst messbar gemacht wird. Sollte dies nicht möglich sein, da der Nutzungskontext möglichst realitätsnah gestaltet werden soll, ist auf Methoden zurückzugreifen, die nicht intrusiv sind und den Nutzer bei der Bewältigung seiner Aufgabe nicht stören. Dies sind entweder qualitative Methoden, die nach Durchführung der Nutzeraufgabe eingesetzt werden können oder Messdaten, die aus einer Simulationsumgebung ausgeleitet werden können sowie Daten, die durch einen externen Beobachter erfasst werden können.

Soll die Cockpit-Funktion jedoch genauer hinsichtlich der genannten Qualitätskriterien untersucht werden, muss die Realitätsnähe entsprechend angepasst und die Messbarkeit der Interaktion des Nutzers mit der Mensch-Maschine-Schnittstelle sichergestellt werden. Das heißt, dass der Nutzungskontext möglichst genau erfasst werden muss. Dies ist beispielsweise anhand der Verwendung einer hierarchischen Task-Analyse sowie einer hierauf aufbauenden Situation-Awareness-Requirements-Analyse möglich. Die Aufgabenanalyse dient hierbei als Grundlage für die Einbindung verschiedener Human-Factors-Methoden, da dieses Vorgehen die Zeitpunkte der Handlungsstrategie des Nutzers kennzeichnet, an welchen Messungen während der Interaktion des Nutzers mit der technischen Funktion möglich sind. Zur Erhebung des Situationsbewusstseins ist eine ergänzende Situation Awareness Requirements Analyse erforderlich, die wiederum kontrastiert, wann der Nutzer welche Informationen benötigt, um zu einer zielführenden Entscheidung zu gelangen. Hierauf aufbauend können insbesondere aufgabenspezifische Methoden wie der SAGAT in die Messumgebung integriert werden. Globale Methoden wie der SART werden hingegen eher retrospektiv angewendet. Beide Formen zur Erhebung des Situationsbewusstseins können durch objektive Messstrategien wie beispielsweise dem Eye-Tracking ergänzt werden. Auch die anderen Qualitätskriterien wie Workload und Gebrauchstauglichkeit können mittels verschiedener subjektiver und objektiver Messmethoden erfasst werden. Bezüglich der Erhebung von Workload bietet sich neben der Anwendung des NASA TLX auch die Umsetzung des Zweitaufgaben-Paradigmas in Form einer DIN-gerechten taktilen DRT an, die ebenfalls durch objektive Eye-Tracking-Maße wie beispielsweise der Pupillen-Dilatation ergänzt werden können. Die Beschreibung der Gebrauchstauglichkeit ist neben den qualitativen Methoden wie Nutzerbefragungen auch anhand der Verwendung von subjektiven Methoden wie dem PSSUQ, aber auch mit objektiven Ansätzen anhand der Erfassung der Bedienzeit möglich.

Zusammenfassung

Um die Eignung der vorgeschlagenen Methoden zur Bestimmung der Qualität der Cockpit-Funktionen bei Bedienung der Nutzerschnittstelle festzustellen, sind diese Methoden zu einem Methoden-Inventar zusammengefügt worden und wurden mit Hilfe von vier verschiedenen Evaluierungen in eine Cockpit-Simulationsumgebung integriert. Zur Eignungsfeststellung des Methoden-Inventars für die Cockpit-Simulationsumgebung ist als Validierungsstrategie ein angepasstes Multi-Trait/ Multi-Method-Vorgehen eingesetzt worden, das ursprünglich von Campbell und Fiske (1959) entwickelt wurde. Die Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit werden hierzu als Konstrukte definiert und lassen sich anhand der verschiedenen Mess-Items beschreiben, die wiederum zu Methoden zusammengefasst werden. Anhand der vier entstehenden, korrelativen Evaluierungsmatrizen kann die konvergente und diskriminante Validität der Methoden und somit deren Eignung festgestellt werden.

Es hat sich gezeigt, dass die eingesetzten Methoden die Qualitätskriterien unterschiedlich gut beschreiben und dass weitgehende Überschneidungen der Konstrukte innerhalb der Evaluierungsmatrizen nachweisbar sind. Die verwendeten subjektiven Methoden können hier für die Bemessung aller Konstrukte beibehalten werden, sollten allerdings nicht als Einzelmethoden eingesetzt werden, da auch hier Überschneidungen aufgezeigt werden konnten. Bei der Verwendung des Eye-Trackings ist insofern Vorsicht geboten, als dass diese Messstrategie nur spezifisch und nicht für alle Konstrukte gleichermaßen eingesetzt werden kann. Dies gilt auch für andere objektive Maße wie beispielsweise die Erfassung der Bedienzeit.

Zudem zeigt sich, dass der Multi-Trait/ Multi-Method-Ansatz als Validierungsstrategie geeignet ist, um die Konstruktvalidität der Methoden im Rahmen der mit Experten erhobenen Evaluierungen zu bestimmen. Allerdings ist auch festzuhalten, dass die Qualität der Validierungsstrategie von der Datenstruktur und der Beschaffenheit des Methoden-Inventars abhängig ist. Allerdings erweist sich dieses Vorgehen hinsichtlich der Abgrenzbarkeit zu anderen Konstrukten wie beispielsweise zwischen Workload und Ermüdung als robust.

Zur Klärung weiterer methodischer Aspekte werden Anpassungen der verwendeten taktilen DRT und eine Überprüfung der Pupillen-Dilatation als Messmethode für Workload empfohlen. Auch eine Anpassung der Matrix unter Verzicht von redundanten Messdaten sowie zusammenfassenden Messwerten wird im Hinblick auf die Verwendung der Validierungsstrategie nahegelegt. Ebenso könnte eine weiterführende, theoriegeleitete Betrachtung der Mess-Items der jeweiligen Konstrukte mit anschließender empirischer Überprüfung zu einer Optimierung des Methodenkonvoluts führen.

Im Rahmen der vorliegenden Arbeit konnte somit gezeigt werden, (i) wie ein Methoden-Inventar für eine Cockpit-Simulationsumgebung zur Bestimmung der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit beschaffen sein soll, (ii) wie dieses in die Simulationsumgebung integriert sowie evaluiert werden kann und (iii) wie dessen Eignung anhand des Multi-Trait/ Multi-Method-Ansatzes festgestellt werden kann. Die verwendeten Methoden wurden zwar insbesondere an die Erhebung in einem Flugzeug-Cockpit-Simulator angepasst; das Methoden-Inventar selbst ist in seiner Grundsubstanz jedoch universell einsetzbar und kann also auch für andere Mensch-Maschine-Schnittstellen in komplexen sozio-technischen Systemen verwendet werden. Grundsätzlich ist her-

Zusammenfassung

auszuheben, dass diese Vorgehensweise der Integration des Methoden-Inventars in eine Simulationsumgebung sowie deren Validierung unter Verwendung von Experten-Stichproben domänenübergreifend anwendbar ist.

LITERATURVERZEICHNIS

- Albert, W. & Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics* (2. Aufl.). Waltham: Morgan Kaufmann.
- Allhoff, D.-W. & Allhoff, W. (2010). *Rhetorik und Kommunikation. Ein Lehr- und Übungsbuch* (15. Aufl.). München: Ernst Reinhardt Verlag.
- Angermeier, D., Christian, B., Bauer, O., Beneken, G., Bergner, K., Birowicz, U. et al. (2006). *V-Modell XT. Das deutsche Referenzmodell für Systementwicklungsprojekte Version: 2.3* (Verein zur Weiterentwicklung des V-Modell XT e.V., Hrsg.). München. Zugriff am 02.04.2020. Verfügbar unter https://www.cio.bund.de/Web/DE/Architekturen-und-Standards/V-Modell-XT/vmodell_xt_node.html
- Annett, J. (2003). Hierarchical Task Analysis. In E. Hollnagel (Hrsg.), *Handbook of cognitive task design* (S. 17–35). Mahwah, New Jersey, London: Lawrence Erlbaum Associates.
- Annett, J. & Duncan, K. D. (1967). *Task analysis and training design*. Report Resumes. Hull University.
- Badke-Schaub, P., Hofinger, G. & Lauche, K. (2008). Human Factors. In P. Badke-Schaub, G. Hofinger & K. Lauche (Hrsg.), *Human Factors: Psychologie sicheren Handelns in Risikobranchen* (S. 1–18). Berlin, Heidelberg: Springer Medizin Verlag.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.
- Banks, V. A., Plant, K. L. & Stanton, N. A. (2018). Development of a Usability Evaluation Framework for the flight deck. In R. Charles & C. Wilkinson (Hrsg.), *Contemporary Ergonomics and Human Factors*. CIEHF.
- Barnum, C. M. (2011). *Usability testing essentials: ready, set-test*. Burlington: Elsevier; Morgan Kaufmann.
- Bartels, M. & Marshall, S. P. (2012). Measuring cognitive workload across different eye tracking hardware platforms. In *Proceedings of the symposium on eye tracking research and applications* (S. 161–164).
- Billeriß, C., Kellerer, J. & Zimmer, A. C. (2018, September). Development, Application, and Validation of a HMI Multimethod Inventory for Cockpit Evaluation - Demonstrated on a newly implemented Navigational Aid. In H. Bercherstorfer, J. Lasry & M. Schwarz (Hrsg.), *Connecting People, Organisations & Technology in Aviation* (S. 509–524). Groningen.
- Bogner, A. & Menz, W. (2002). Die methodische Mehrdeutigkeit des Experteninterviews. In A. Bogner, B. Littig, W. Menz & others (Hrsg.), *Das Experteninterview. Theorie, Methode, Anwendung* (2. Aufl., S. 33–70). Wiesbaden: Springer Fachmedien.
- Bojko, A. (2013). *Eye tracking the user experience. A practical guide to research*. Brooklyn, New York: Rosenfeld Media.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bowman, D. S. & Hanowski, R. J. (2008). PERCLOS+: Development of a robust field measure of driver drowsiness. In ITSA (Hrsg.), *ITS America's 2008 Annual Meeting* (S. 1–12).

- Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland & B. Weerdmeester (Hrsg.), *Usability Evaluation In Industry* (S. 189–194). CRC Press. Zugriff am 18.03.20. Verfügbar unter google books
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Byers, J. C., Bittner, A. C. & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary. In A. Mital (Hrsg.), *Advances in Industrial Ergonomics and Safety I*. London: Taylor & Francis.
- Cain, B. (2007). *A review of the mental workload literature* (Defence Research And Development Toronto (Canada), Hrsg.). Toronto: Human System Integration Section.
- Caldwell, J. A., Wilson, G. F., Cetinguc, M., Gaillard, A. W., Gunder, A., Lagarde, D. et al. (1994). *Psychophysiological Assessment Methods sur Microfiches* (North Atlantic Treaty Organization, Hrsg.) (AGARD-AR-324). Neuilly sur Seine: Advisory Group for Aerospace Research & Development.
- Calhoun, G. L. (2000). *Pilot vehicle interface*. RTO-EN-12 (Research and Technology Organization (RTO) of NATO, Hrsg.). Neuilly sur Seine Cedex.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, 56(2), 81–105. Zugriff am 12.11.2016.
- Clamann, M. & Kaber, D. B. (2004). Applicability of usability evaluation techniques to aviation systems. *The international journal of aviation psychology*, 14(4), 395–420.
- Cockburn, A. (2001). *Writing effective use cases*: Addison-Wesley Professional.
- Coombs, L. F. E. (1999). *Fighting Cockpits 1914 - 2000. Design and development of military aircraft cockpits*. Shrewsbury, England: Airlife Publishing Ltd.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin*, 52(4), 281. Zugriff am 29.12.2019. Verfügbar unter <https://psycnet.apa.org/record/1956-03730-001>
- Czaja, S. J. & Sankaran, N. N. (2012). Human Factors Engineering and Systems Design. In G. Salvendy (Hrsg.), *Handbook of human factors and ergonomics* (4. edition, S. 38–57). Hoboken, NJ: Wiley.
- Demberg, V. (2013). Pupillometry: the index of cognitive activity in a dual-task study. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Bd. 35, S. 2154–2159).
- Diekmann, A. (2005). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen* (14. Aufl.). Reinbek: Rowohlt Taschenbuch Verlag.
- Döhl, M. (2001). Walkthrough vs. Videokonfrontation Vergleich zweier Methoden zur formativen Software-Evaluation. In *Mensch & Computer 2001. 1. Fachübergreifende Konferenz* (S. 441–442).
- Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D. & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly*, 6(1), 1–20.
- Durso, F. T. & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors*, 50(3), 442–448.

- Eichinger, A. (2011). *Bewertung von Benutzerschnittstellen für Cockpits hochagiler Flugzeuge*. Dissertationsschrift. Universität Regensburg, Regensburg.
- Ellgring, H. (1995). Audiovisuell unterstützte Beobachtung. In U. Flick, E. von Kardorff, H. Keupp, L. von Rosenstiel & S. Wolff (Hrsg.), *Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen* (2. Aufl., S. 203–208). Weinheim: Beltz Psychologie Verlags Union.
- Endsley, M. R. (1988a). Design and Evaluation for Situation Awareness Enhancement. In Human Factors Society (Hrsg.), *Proceedings of the Human Factors Society* (S. 97–101). Zugriff am 30.01.2020.
- Endsley, M. R. (1988b). Situation awareness global assessment technique (SAGAT). In IEEE (Hrsg.), *Proceedings of the IEEE 1988 national aerospace and electronics conference* (S. 789–795).
- Endsley, M. R. (1990). A Methodology for the Objective Measurement of Pilot Situation Awareness. In Advisory Group for Aerospace Research & Development (Hrsg.), *AGARD Conference Proceedings No. 478* (Bd. 478, 1-1 - 1-9). Neuilly sur Seine: AGARD.
- Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *The international journal of aviation psychology*, 3(2), 157–168.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Endsley, M. R. (1999). Situation Awareness in Aviation Systems. In D. J. Garland, J. A. Wise & V. D. Hopkin (Hrsg.), *Handbook of Aviation Human Factors* (S. 257–276). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Endsley, M. R. (2000a). Direct Measurement of Situation Awareness: Validity and Use of SAGAT. In M. R. Endsley & D. J. Garland (Hrsg.), *Situation awareness analysis and measurement* (S. 147–173). New York: CRC Press.
- Endsley, M. R. (2000b). Theoretical Underpinnings of Situation Awareness: A Critical Review. In M. R. Endsley & D. J. Garland (Hrsg.), *Situation awareness analysis and measurement* (S. 3–32). New York: CRC Press.
- Endsley, M. R. (2001). Designing for situation awareness in complex systems. In *Proceedings of the Second International Workshop on symbiosis of humans, artifacts and environment* (S. 1–14).
- Endsley, M. R. & Jones, D. G. (2012). *Designing for situation awareness: An approach to user-centered design*: CRC Press.
- Endsley, M. R., Selcon, S. J., Hardiman, T. D. & Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting* (Bd. 42, S. 82–86).
- Endsley, M. R., Sollenberger, R. & Stein, E. (2000). Situation Awareness: A Comparison of Measures. In SA Technologies (Hrsg.), *Proceedings of the Human Performance, Situation Awareness and Automation* (S. 15–19). Savannah, GA: SA Technologies.
- ISO 9241-11:2018(E) (März 2018). *Ergonomics of human-system interaction*. Genf: ISO 2018.
- Technical Report, ISO/TR16982:2002(E) (2002). *Ergonomics of human-system-interaction - Usability methods supporting human-centred design*. Genf: ISO 2002.

- DIN EN ISO 9241-210 (Januar 2011). *Ergonomie der Mensch-System-Interaktion*. Berlin: Beuth Verlag GmbH.
- EN ISO 10075-1 (Juli 2000). *Ergonomic principles related to mental workload*: CEN.
- Fracker, M. L. (1988). A Theory of Situation Assessment: Implications for Measuring Situation Awareness. In Human Factors Society (Hrsg.), *Proceedings of the Human Factors Society* (S. 102–106).
- Fracker, M. L. (1991). *Measures of situation awareness: Review and future directions* (Airforce Systems Command, Hrsg.). Ohio: Crew Systems Directorate Human Engineering Division.
- Fruhling, A. & Lee, S. (2005). Assessing the reliability, validity and adaptability of PSSUQ. In Association for Information Systems (Hrsg.), *Americas Conference on Information Systems (AMCIS) 2005 Proceedings* (S. 2394–2402). AIS Electronic Library (AISeL). Zugriff am 29.05.2020.
- Gartner, W. B. & Murphy, M. R. (1976, November). *Pilot workload and fatigue: A critical survey of concepts and assessment techniques* (National Aeronautics and Space Administration, Hrsg.). Los Gatos: Nasa Ames Research Centre.
- Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook* (2nd ed.). Boca Raton: CRC Press.
- Gediga, G., Greif, S., Monecke, U. & Hamborg, K.-C. (1989). Aufgaben-und Tätigkeitsanalysen als Grundlage der Softwaregestaltung. *Software-Ergonomie'89: Aufgabenorientierte Systemgestaltung und Funktionalität*.
- Goldberg, J. H. & Kotval, X. P. (1998). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631–645.
- Goldberg, J. H. & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In R. Godijn, J. Theeuwes, J. Hyona & R. Radach (Hrsg.), *The Mind's Eye: Cognitive and applied aspects of eye movement* (S. 493–516). North Holland: Elsevier.
- Goldstein, B. E. (1999). *Sensation & Perception* (5. Aufl.). Pacific Grove: Brooks/Cole Publishing Company.
- Gopher, D. & Donchin, E. (1986). Workload - an examination of the concept. In K. R. Boff, L. Kaufman & J. P. Thomas (Hrsg.), *Handbook of Perception and Human Performance. Volume 2. Cognitive Processes and Performance*. (41-1 - 41-49). New York: John Wiley & Sons.
- Grandt, M. (2004). *Zur Erfassung und Bewertung der mentalen Beanspruchung mittels psychophysiologischer Messverfahren*. Bergische Universität Wuppertal, Wuppertal.
- Hanington, B. & Martin, B. (2012). *Universal Methods of Design. 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Beverly: Rockport Publishers.
- Harbluk, J. L., Burns, P. C., Hernandez, S., Tam, J. & Glazduri, V. (2013). Detection Response Tasks: Using Remote, Headmounted and Tactile Signals to Assess Cognitive Demand While Driving. In University of Iowa (Hrsg.), *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (S. 78–84).
- Harris, R. L., Tole, S. R., Stephens, A. T. & Ephrath, A. R. (1981). Visual Scanning Behavior and Pilot Workload. In Nasa STI Facility (Hrsg.), *Symposium on Aviation Psychology* (S. 1–9).

- Hart, S. G. (Hrsg.) (2006). *NASA-task load index (NASA-TLX); 20 years later* (Bd. 50). Los Angeles, CA: Sage Publications CA.
- Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Bd. 52, 139-183).
- Hart, S. G. & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Hrsg.), *Manprint* (S. 257–296). Dordrecht: Springer.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 144–171). Berlin, Heidelberg: Springer.
- Hendy, K. C., Hamilton, K. M. & Landry, L. N. (1993). Measuring Subjective Workload: When Is One Scale Better Than Many? *Human Factors*, 35(4), 579–601.
- Hering, E. (2014). Methoden des Zeitmanagements. In E. Hering (Hrsg.), *Zeitmanagement für Ingenieure* (S. 13–27). Wiesbaden: Springer-Verlag.
- Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner Jr, A. C., Zaklade, A. L. & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors*, 34(4), 429–439.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: OUP Oxford.
- Hopf, C. (1995). Qualitative Interviews in der Sozialforschung. Ein Überblick. In U. Flick, E. von Kardorff, H. Keupp, L. von Rosenstiel & S. Wolff (Hrsg.), *Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen* (2. Aufl., S. 177–181). Weinheim: Beltz Psychologie Verlags Union.
- Institute of Electrical and Electronics Engineers. (1990). *IEEE standard glossary of software engineering terminology: Approved September 28, 1990, IEEE Standards Board* (Revision and redesignation of IEEE Std 792-1983). New York, NY: Inst. of Electrical and Electronics Engineers.
- Johnson, G. I., Clegg, C. W. & Ravden, S. J. (1989). Towards a practical method of user interface evaluation. *Applied Ergonomics*, 20(4), 255–260.
- Johnson, G. I. (1996). The usability checklist approach revisited. In P. W. Jordan, B. Thomas, I. L. McClelland & B. A. Weerdmeester (Hrsg.), *Usability evaluation in industry* (S. 179–188). London: Taylor & Francis. Zugriff am 02.06.2020.
- Jones, D. G. (2000). Subjective Measures of Situation Awareness. In M. R. Endsley & D. J. Garland (Hrsg.), *Situation awareness analysis and measurement* (S. 113–128). New York: CRC Press.
- Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329–354.
- Kahneman, D. (1973). *Attention and effort* (Bd. 1063). New Jersey: Prentice-Hall, Inc.
- Kantowitz, B. H. & Casper, P. A. (1988). Human Workload in Aviation. In E. L. Wiener & D. C. Nagel (Hrsg.), *Human factors in aviation*. San Diego: Academic Press Inc.
- Karwowski, W. (2001). *International Encyclopedia of Ergonomics and Human Factors* (Volume 1). New York: Taylor & Francis.

- Karwowski, W. (2012). The Discipline of Human Factors and Ergonomics. In G. Salvendy (Hrsg.), *Handbook of human factors and ergonomics* (4. edition, S. 3–37). Hoboken, NJ: Wiley.
- Kellerer, J. P. (2010). *Untersuchung zur Auswahl von Eingabeelementen für Großflächendisplays in Flugzeugcockpits*. Dissertation. Technische Universität Darmstadt, Darmstadt.
- Kokini, C. M., Lee, S., Richard, K. J. & Moon, S. K. (2012). Considering Context: The Role of Mental Workload and Operator Control in Users' Perceptions of Usability, *28*, 543–559.
- Kramer, A. F. (1990). *Physiological metrics of mental workload: A review of recent progress* (Navy Personnel Research and Development center, Hrsg.). San Diego: University of Illinois at Urbana-Champaign.
- Krems, J. F. (1997). Expertise und diagnostische Urteilsbildung. In K. Sonntag & N. Schaper (Hrsg.), *Störungsmanagement und Diagnosekompetenz. Leistungskritisches Denken und Handeln in komplexen technischen Systemen* (S. 77–94). Zürich: vdf Hochschulverlag AG.
- Le Blaye, P. (2000). *Agility: History, Definitions and Basic Concepts*. RTO-EN-12 (Research and Technology Organization (RTO) of NATO, Hrsg.). Neuilly sur Seine Cedex.
- Lewis, C., Polson, P. G., Wharton, C. & Rieman, J. (1990). Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In CHI'90 (Hrsg.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (S. 235–242). Zugriff am 31.05.2020.
- Lewis, J. R. (1992). Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Bd. 36, S. 1259–1260). Los Angeles, CA: Sage Publications CA.
- Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction*, *14*(3-4), 463–488.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz Psychologie Verlags Union.
- Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D. & Linton, P. M. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies* (U.S. Army Research Institute for the Behavioral and Social Sciences, Hrsg.). Systems Research Laboratory.
- Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In Rösler Frank (Hrsg.), *Sonderdruck aus Enzyklopädie der Psychologie. Ergebnisse und Anwendungen der Psychophysiologie* (Bd. 100, S. 799–864). Göttingen: Hogrefe.
- Marshall, S. P. (1999), 6, 090, 051.
- Marshall, S. P., Pleydell-Pearce, C. W. & Dickson, B. T. (2003). Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. In HICSS 36 (Hrsg.), *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, 2003*. (6-pp).
- McGuiness, B. & Ebbage, L. (2002). *Assessing Human Factors in Command and Control: Workload and Situational Awareness Metrics* (BAE Systems Advanced Technology Centre, Hrsg.). Bristol.
- McNicol, D. (2005). *A primer of signal detection theory*: Psychology Press.

- Megaw, T. (2005). The definition and measurement of mental workload. In J. R. Wilson & E. N. Corlett (Hrsg.), *Evaluation of Human Work* (S. 525–552). Boca Raton: CRC Press. Zugriff am 11.05.2020.
- Messick, S. (1987). Validity. *ETS Research Report Series, 1987(2)*, i-208.
- DIN EN ISO 15007-1:2003-10 (Oktober 2003). *Messung des Blickverhaltens von Fahrern bei Fahrzeugen mit Fahrinformations- und Assistenzsystemen*. Berlin: Beuth Verlag GmbH.
- Moore, K. S. (2009). *Comparison of eye movement data to direct measures of situation awareness for development of a novel measurement technique in dynamic, uncontrolled test environments*. Dissertation. Graduate School of Clemson University. Zugriff am 05.06.2020.
- Moore, K. S. & Gugerty, L. (2010). Development of a novel measure of situation awareness: The case for eye movement analysis. In *Proceedings of the human factors and ergonomics society annual meeting* (Bd. 54, S. 1650–1654). Los Angeles, CA: Sage Publications.
- Morgenstern, C., Kellerer, J., Neujahr, H. & Zimmer, A. (2016). Multidimensionales Methodeninventar zur Überprüfung der kognitiven Ergonomie eines Flugzeugcockpits. In C. Wienrich, T. Zander & K. Gramann (Hrsg.), *Trends in Neuroergonomics: Tagungsband 11. Berliner Werkstatt Mensch-Maschine-Systeme; 7.-9. Oktober 2015* (S. 148–153). Berlin: Universitätsverlag der TU Berlin.
- Muehlethaler, C. M. & Knecht, C. P. (2016). Situation awareness training for general aviation pilots using eye tracking. *IFAC-PapersOnLine, 49(19)*, 66–71.
- Neisser, U. (1978). Perceiving, anticipating, and imagining.
- Nielsen, J. (1994). Usability inspection methods. In Conference Companion (Hrsg.), *Conference companion on Human factors in computing systems* (S. 413–414).
- Nielsen, J. & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (S. 206–213). Zugriff am 18.04.20.
- Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327(1241)*, 585–593.
- Norman, D. A. (2002). Emotion & Design: Attractive Things Work Better. *Interactions Magazine, 9*, 36–42.
- Nygren, T. E. (1991). Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload. *Human Factors, 33(1)*, 17–33.
- O'Donnell, R. D. & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman & J. P. Thomas (Hrsg.), *Handbook of Perception and Human Performance. Volume 2. Cognitive Processes and Performance*. (42-1 - 42-49). New York: John Wiley & Sons.
- Peißl, S., Wickens, C. D. & Baruah, R. (2018). Eye-tracking measures in aviation: a selective literature review. *The International Journal of Aerospace Psychology, 28(3-4)*, 98–112.
- Perrot, A., Schader, N. T., Leonhardt, J. & Licu, T. (2019). *Human Factors Integration in ATM System Design. A white Paper* (Eurocontrol, Hrsg.).

- Pew, R. W. (2000). The State of Situation Awareness Measurement. Heading toward the next century. In M. R. Endsley & D. J. Garland (Hrsg.), *Situation awareness analysis and measurement* (S. 33–50). New York: CRC Press.
- Pradhan, A. K., Fisher, D. L. & Pollatsek, A. (2006). Risk Perception Training for Novice Drivers. Evaluating Duration of Effects of Training on a Driving Simulator. *Journal of the Transportation Research Board*, (1969), 58–64.
- Quesenbery, W. (2004). Balancing the 5Es of usability. *Cutter IT Journal*, 17(2), 4–11.
- Rabl, A. N. (2016). *Bewertung räumlicher Anzeigen im Flugzeugcockpit*. Dissertation. Universität Regensburg, Regensburg.
- Rabl, A. N., Neujahr, H. & Zimmer, A. (2016). Evaluation of spatial audio in a fighter cockpit. In EAAP (Hrsg.), *Proceedings of the 32nd EAAP Conference* (S. 26–30).
- Rauch, N. (2009). *Ein verhaltensbasiertes Messmodell zur Erfassung von Situationsbewusstsein im Fahrkontext*. Dissertation. Universität Würzburg, Würzburg. Zugriff am 22.05.2020.
- Ravden, S. & Johnson, G. (1989). *Evaluating usability of human-computer interfaces: a practical method*. Chichester: Halsted Press.
- Technical Specification, ISO/TS 15007-2:2001(E) (Dezember 2001). *Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems*. Genf: ISO 2001.
- ISO 17488:2016 (Oktober 2016). *Road Vehicles -Transport Information and Control Systems*. Genf: ISO 2016.
- Robinski, M. & Stein, M. (2013). Tracking visual scanning techniques in training simulation for helicopter landing. *Journal for Eye Movement Research*, 6(2), 1–17.
- Rupp, C. (2014). In medias RE. In C. Rupp (Hrsg.), *Requirements-Engineering und -Management: Aus der Praxis von klassisch bis agil* (6., aktual. u. erw. Aufl., S. 9–30). München: Carl Hanser Verlag.
- Salmon, P. M., Stanton, N. A., Jenkins, D. P. & Walker, G. H. (2009). *Distributed situation awareness. Theory, measurement and application to teamwork*. Farnham: Ashgate Publishing Limited.
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D. P., Ladva, D. & Rafferty, L. (2007). *Measuring Situation Awareness in Command and Control: Comparison of Methods Study* (2. Aufl.): Human Factors Integration Defence Technology Centre. Zugriff am 04.02.2020.
- Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L. et al. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, 39(3), 490–500. Zugriff am 12.06.2020.
- Sarter, N. B. & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The international journal of aviation psychology*, 1(1), 45–57.
- Sauro, J. & Lewis, J. R. (2009). Correlations among prototypical usability metrics: evidence for the construct of usability. In Association for Computing Machinery (Hrsg.), *Proceedings of the SIGCHI conference on human factors in computing systems* (S. 1609–1618). New York. Zugriff am 28.05.2020.

- Sauro, J. & Lewis, J. R. (2012). *Quantifying the user experience: Practical statistics for user research*: Morgan Kaufmann.
- Schmermelleh-Engel, K. & Schweizer, K. (2012). Multitrait-Multimethod-Analysen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 346–361). Berlin, Heidelberg: Springer.
- Schneider, W. & Detweiler, M. (1988). The role of practice in dual-task performance: toward workload modeling a connectionist/control architecture. *Human Factors*, 30(5), 539–566.
- Schnell, R., Hill, P. B. & Esser, E. (2018). *Methoden der empirischen Sozialforschung* (11. Aufl.). Berlin: Walter de Gruyter.
- Schwalm, M. (2009). *Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext*. Dissertation. Universität des Saarlandes, Saarbrücken.
- Selcon, S. J. & Taylor, R. M. (1996). Evaluation of the Situational Awareness Rating Technique (SART) as a tool for aircrew systems design. In Advisory Group for Aerospace Research & Development (Hrsg.), *AGARD Conference Proceedings No. 575* (S. 23–53). Neuilly sur Seine: AGARD.
- Selcon, S. J., Taylor, R. M. & Koritsas, E. (1991). Workload or Situational Awareness?: TLX vs. SART for Aerospace Systems Design Evaluation. In Human Factors Society (Hrsg.), *Proceedings of the Human Factors Society Annual Meeting* (Bd. 2, S. 62–66). Los Angeles: Sage Publications.
- Shepherd, A. (2005). *Hierarchical task analysis*. London: Taylor & Francis.
- Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N. & Diakopoulos, N. (2005). *Designing the user interface. Strategies for effective human-computer interaction* (4. Aufl.). Boston: Pearson.
- Shorrock, S., Leonhardt, J., Licu, T. & Peters, C. (2014). *Systems Thinking for Safety: Ten Principles. Moving towards Safety -II*. A White Paper (Eurocontrol, Hrsg.). Zugriff am Mai 2020.
- Smith, K. & Hancock, P. A. (1994). Situation Awareness is Adaptive, Externally-Directed Consciousness. In R. D. Gilson, D. J. Garland & J. M. Koonce (Hrsg.), *Proceedings of the Center for Applied Human Factors in Aviation (CAHFA) Conference on Situational Awareness in Complex Systems* (S. 59–68). Daytona Beach, Florida USA: Embry-Riddle Aeronautical University Press.
- Sommer, D. & Golz, M. (2010). Evaluation of PERCLOS based Current Fatigue Monitoring Technologies. In IEEE Engineering in Medicine and Biology (Hrsg.), *Annual International Conference of the IEEE* (S. 4456–4459).
- Sonderegger, A. & Sauer, J. (2009). The influence of laboratory set-up in usability tests: effects on user performance, subjective ratings and physiological measures. *Ergonomics*, 11(52), 1350–1361.
- Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C. & Daniel, J. P. (2011). *Human Factors Methods. A Practical Guide for Engineering and Design* (4. Aufl.). Surrey, Burlington: Ashgate Publishing Limited.
- Taylor, R. M. (1990). Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design. In Advisory Group for Aerospace Research & Development (Hrsg.), *AGARD Conference Proceedings No. 478* (Bd. 478, 3-1 - 3-17). Neuilly sur Seine: AGARD. Zugriff am 29.01.2020.

- Turner, C. W., Lewis, J. R. & Nielsen, J. (2006). Determining usability test sample size. *International encyclopedia of ergonomics and human factors*, 3(2), 3084–3088.
- Turner, J. (1997). Cockpit Usability - A Design Checklist. In Advisory Group for Aerospace Research & Development (Hrsg.), *AGARD Conference Proceedings 600 (A22-1 - A22-7)*. Neuilly sur Seine: AGARD.
- Uhlarik, J. & Comerford, D. A. (2002). *A review of situation awareness literature relevant to pilot surveillance functions* (Federal Aviation Administration, Hrsg.). Washington: Federal Aviation Administration.
- Van Dijk, H., van de Merwe, K. & Zon, R. (2011). A coherent impression of the pilots' situation awareness: Studying relevant human factors tools. *The international journal of aviation psychology*, 21(4), 343–356.
- Vidulich, M. A. (2003). Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice. In P. S. Tsang & M. A. Vidulich (Hrsg.), *Principles and practice of aviation psychology* (S. 115–146). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Vidulich, M. A. & Tsang, P. S. (1987). Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. In *Proceedings of the Human Factors Society Annual Meeting* (Bd. 31, S. 1057–1061).
- Vidulich, M. A. & Tsang, P. S. (2012). Mental Workload and Situation Awareness. In G. Salvendy (Hrsg.), *Handbook of human factors and ergonomics* (4. edition, S. 243–264). Hoboken, NJ: Wiley.
- Wickens, C. D. (2003). Aviation Displays. In P. S. Tsang & M. A. Vidulich (Hrsg.), *Principles and practice of aviation psychology* (S. 147–200). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wickens, C. D., Gordon, S. E. & Liu, Y. (1998). *An introduction to human factors engineering*. New York: Longman.
- Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (Eds.) (2013). *Engineering psychology and human performance* (Always learning, 4 // 4. ed., internat. ed.). Boston: Pearson.
- Wierwille, W. W., Ellsworth, L. A., Wreggit, S. S., Kirn, C. L. & Fairbanks, R. J. (1994). *Research on Vehicle-Based Driver Status / Performance Monitoring. Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness* (Office of Crash Avoidance Research - National Highway Traffic Safety Administration, Hrsg.). Blacksburg: Vehicle Analysis and Simulation Laboratory.
- Wierwille, W. W., Williges, R. C. & Schiflett, S. G. (1979). Aircrew workload assessment techniques. *Survey of methods to assess workload*, 19–53.
- Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors*, 21(5), 575–593.
- Wilkinson, P. R. (1992). The Integration of Advanced Cockpit and Systems Design. In North Atlantic Treaty Organization (Hrsg.), *AGARD Conference Proceedings 521 (26 / 1-9)*. Neuilly sur Seine: AGARD-CP-521.
- Williges, R. C., Williges, B. H. & Fainter, R. G. (1988). Software Interfaces for Aviation Systems. In E. L. Wiener & D. C. Nagel (Hrsg.), *Human factors in aviation* (S. 463–494). San Diego: Academic Press Inc.

Literaturverzeichnis

- Williges, R. C. & Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors*, 21(5), 549–574.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Verlag Hans Huber.
- Yeh, Y. H. & Wickens, C. D. (1984). *The dissociation of subjective measures of mental workload and performance. Final Report* (National Aeronautics and Space Administration, Hrsg.). Urbana - Champaign: Engineering - Psychology Research Laboratory University of Illinois.
- Young, R. A., Hsieh, L. & Seaman, S. (2013). The Tactile Detection Response Task: Preliminary validation for measuring the attentional effects of cognitive load. In University of Iowa (Hrsg.), *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (S. 71–77).
- Zimmer, A. C. (2016). The Systemic Character of Perception and Action in Cockpit Design. In EAAP (Hrsg.), *Proceedings of the 32nd EAAP Conference*.

ABKÜRZUNGSVERZEICHNIS

AOI	Area of Interest
ATC	Air Traffic Controller, Fluglotse
3DA-Evaluierung	Evaluierung eines räumlich-auditiven Warnsignals
GT	Gebrauchstauglichkeit
HMI	Human Machine Interface
HMMT-Matrix	Hetero-Method-Mono-Trait-Matrix
HMHT-Matrix	Hetero-Method-Hetero-Trait-Matrix
HTMM	Hetero-Trait-Mono-Method
HTHM	Hetero-Trait-Hetero-Method
HTA	Hierarchische Task-Analyse
MMI	Mensch-Maschine-Interaktion
Mdn	Median
MMS	Mensch-Maschine-Schnittstelle
MS	Machbarkeitsstudie
MTHM-Diagonale	Mono-Trait-Hetero-Method-Diagonale
MTMM-Ansatz	Multi-Trait/ Multi-Method-Ansatz
MTMM-Matrix	Multi-Trait/ Multi-Method-Matrix
MTMM-Diagonale	Mono-Trait-Mono-Method-Diagonale
MMMT-Matrix	Mono-Method-Mono-Trait-Matrix
NASA TLX	NASA Task Load Index
RD-Evaluierung	Re-Design-Evaluierung
SAGAT	Situation Awareness Global Assessment Technique
SARA	Situation Awareness Requirements Analyse
SART	Situation Awareness Rating Technique
SB	Situationsbewusstsein
SEN-Evaluierung	Evaluierung der Bedienung zur Einstellung von Sensitivität und Genauigkeit eines Sensors
SPAM	Situation Awareness Present Assesment Method
VP	Versuchsperson
WL	Workload

ABBILDUNGSVERZEICHNIS

Abbildung 1.1 Vorgehen zur Implementierung und Eignungsfeststellung eines Methoden-Inventars	4
Abbildung 2.1 Darstellung des Cockpit-Simulators	11
Abbildung 4.1 Exemplarischer HTA-Plan der Aufgabe ‚Flugzeugsteuerung‘ mit integrierter SARA nach Morgenstern et al. (2016)	49
Abbildung 5.1 Boxplots des 3D SART nach dem 1. Szenario	60
Abbildung 5.2 Boxplots des 3D SART nach dem 2. Szenario	60
Abbildung 5.3 Boxplots der 3D SART-Gesamtwerts (Situationsbewusstsein) in Abhängigkeit der Szenario-Reihenfolge	61
Abbildung 5.4 SAGAT-Reizmaterial der stummen Karte mit Fixpunkten für das Szenario München.....	62
Abbildung 5.5 SAGAT-Lösung der stummen Karte für das Szenario München.....	62
Abbildung 5.6 Boxplots der SAGAT-Antwortleistung in Abhängigkeit der Szenario-Reihenfolge	63
Abbildung 5.7 Boxplots des SAGAT-Sicherheitsurteils in Abhängigkeit der Szenario-Reihenfolge	63
Abbildung 5.8 Boxplots der SAGAT-Items in Abhängigkeit der Szenario-Reihenfolge	64
Abbildung 5.9 Boxplots der mittels SAGAT erinnerten Navigationspunkte in Abhängigkeit der Szenario-Reihenfolge	65
Abbildung 5.10 Boxplots der SAGAT-Sicherheitsurteile bezüglich der Navigationspunkte in Abhängigkeit der Szenario-Reihenfolge	65
Abbildung 5.11 Boxplots der relativen Passung der Probandenantworten in der Sortieraufgabe in Abhängigkeit der Szenario-Reihenfolge	66
Abbildung 5.12 Boxplots der relativen Passung der Probandenantworten zur Ideallösung in der Sortieraufgabe in Abhängigkeit der Szenario-Reihenfolge.....	66
Abbildung 5.13 Anzahl der Fixationen während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge	67
Abbildung 5.14 Fixationsdauer während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge.....	67
Abbildung 5.15 Anzahl der Fixationen während der Szenarien in Abhängigkeit der Szenario-Reihenfolge	68
Abbildung 5.16 Fixationsdauer während der Szenarien in Abhängigkeit der Szenario-Reihenfolge....	68
Abbildung 5.17 NASA TLX Boxplots der Workload-Dimensionen des ersten Szenarios	69
Abbildung 5.18 NASA TLX Boxplots der Workload-Dimensionen des zweiten Szenarios.....	70
Abbildung 5.19 Boxplots des NASA TLX Gesamt-Workloads in Abhängigkeit der Szenario-Reihenfolge	71
Abbildung 5.20 Boxplots der TDRT-Reaktionszeiten in Abhängigkeit der Szenario-Reihenfolge mit Darstellung der Baseline	72

Abbildungsverzeichnis

Abbildung 5.21 Boxplots der TDRT-Hit-Rates in Abhängigkeit der Szenario-Reihenfolge mit Darstellung der Baseline	72
Abbildung 5.22 Boxplots der Pupillenerweiterung in Abhängigkeit der Szenario-Reihenfolge.....	73
Abbildung 5.23 Boxplots des PERCLOS in Abhängigkeit der Szenario-Reihenfolge	73
Abbildung 5.24 Boxplots der Eingabezeit in Abhängigkeit der Szenario-Reihenfolge	75
Abbildung 5.25 Boxplots der PSSUQ-Faktoren.....	76
Abbildung 5.26 Gegenüberstellung der PSSUQ-Mediane der Machbarkeitsstudie (MS) und der Re-Design-Studie (RD).....	76
Abbildung 5.27 Boxplots der prozentualen Transitionszeit in Abhängigkeit der Szenario-Reihenfolge	77
Abbildung 5.28 Boxplots der Anzahl an Sakkaden in Abhängigkeit der Szenario-Reihenfolge	78
Abbildung 5.29 Boxplots der Sakkaden-Dauer in Abhängigkeit der Szenario-Reihenfolge	78
Abbildung 5.30 Boxplots der Anzahl an Sakkaden während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge	78
Abbildung 5.31 Boxplots der Sakkaden-Dauer während der Nutzung der Navigationshilfe in Abhängigkeit der Szenario-Reihenfolge	78
Abbildung 6.1 Validitätsformen und deren Teilaspekte nach Bühner (2011).....	89
Abbildung 6.2 Exemplarische Multi-Trait/ Multi-Method-Matrix angepasst nach Campbell und Fiske (1959).....	92
Abbildung 6.3 Schematische Darstellung einer Evaluationsmatrix.....	94
Abbildung 6.4 Schematische Illustration einer MMT-Matrix für das Qualitätskriterium Situationsbewusstsein	95
Abbildung 6.5 Schematische Illustration einer HMMT-Matrix für das Qualitätskriterium Situationsbewusstsein	96
Abbildung 6.6 Schematische Illustration der HMHT-Matrizen für die Kombinationen Situationsbewusstsein mit Workload (SB-WL) und Situationsbewusstsein mit Gebrauchstauglichkeit (SB-GT)	97
Abbildung 7.1 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Erfassung des Situationsbewusstseins	100
Abbildung 7.2 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Erfassung des Situationsbewusstseins	101
Abbildung 7.3 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Workload	101
Abbildung 7.4 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Gebrauchstauglichkeit.....	102
Abbildung 7.5 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Workload	103

Abbildungsverzeichnis

Abbildung 7.6 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Workload	104
Abbildung 7.7 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Situationsbewusstsein	105
Abbildung 7.8 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Gebrauchstauglichkeit	106
Abbildung 7.9 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit	106
Abbildung 7.10 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit	107
Abbildung 7.11 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Situationsbewusstsein	108
Abbildung 7.12 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Workload	109
Abbildung 7.13 Anteil an bedeutsamen Korrelationen in den MMT-Matrizen über alle Evaluierungen	109
Abbildung 7.14 Gegenüberstellung der konvergenten und der diskriminanten Validität anhand des Anteils an bedeutsamen Korrelationen über alle Evaluierungen	110
Abbildung 7.15 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des 3D SARTs	112
Abbildung 7.16 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des 3D SARTs	112
Abbildung 7.17 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Workload	113
Abbildung 7.18 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Gebrauchstauglichkeit	113
Abbildung 7.19 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des SAGATs	114
Abbildung 7.20 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs	114
Abbildung 7.21 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Workload	115
Abbildung 7.22 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Gebrauchstauglichkeit	115
Abbildung 7.23 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Sortieraufgabe	115
Abbildung 7.24 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Gefahren-Detektierung	116
Abbildung 7.25 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit	117

Abbildungsverzeichnis

Abbildung 7.26	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit	117
Abbildung 7.27	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Workload	118
Abbildung 7.28	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit	118
Abbildung 7.29	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit	119
Abbildung 7.30	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit	119
Abbildung 7.31	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Workload	120
Abbildung 7.32	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit	120
Abbildung 7.33	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des NASA TLX.....	121
Abbildung 7.34	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des NASA TLX.....	121
Abbildung 7.35	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Situationsbewusstsein.....	122
Abbildung 7.36	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Gebrauchstauglichkeit.....	122
Abbildung 7.37	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der TDRT.....	123
Abbildung 7.38	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der TDRT.....	123
Abbildung 7.39	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Situationsbewusstsein	124
Abbildung 7.40	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Gebrauchstauglichkeit	124
Abbildung 7.41	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung	125
Abbildung 7.42	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung	125
Abbildung 7.43	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Situationsbewusstsein	126
Abbildung 7.44	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Gebrauchstauglichkeit	126
Abbildung 7.45	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung PERCLOS	127

Abbildungsverzeichnis

Abbildung 7.46	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung PERCLOS	127
Abbildung 7.47	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Situationsbewusstsein	128
Abbildung 7.48	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Gebrauchstauglichkeit	128
Abbildung 7.49	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung	129
Abbildung 7.50	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung	129
Abbildung 7.51	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Situationsbewusstsein	130
Abbildung 7.52	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Gebrauchstauglichkeit	130
Abbildung 7.53	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Zeitleistung	131
Abbildung 7.54	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Zeitleistung	131
Abbildung 7.55	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Reaktionszeiten in Kombination mit Situationsbewusstsein	132
Abbildung 7.56	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Reaktionszeiten in Kombination mit Workload.....	132
Abbildung 7.57	Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Geschwindigkeitsleistung in der 3DA-Studie.....	133
Abbildung 7.58	Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Winkelgenauigkeit (Notch) in der 3DA-Studie.....	134
Abbildung 7.59	Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Steuereingaben in der 3DA-Studie	135
Abbildung 7.60	Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Genauigkeitsmessung Azimut in der 3DA-Studie	136
Abbildung 7.61	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des PSSUQ	136
Abbildung 7.62	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des PSSUQ	136
Abbildung 7.63	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Situationsbewusstsein	137
Abbildung 7.64	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Workload	137
Abbildung 7.65	Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking Zeitleistung.....	138

Abbildungsverzeichnis

Abbildung 7.66 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking Zeitleistung.....	138
Abbildung 7.67 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Situationsbewusstsein.....	139
Abbildung 7.68 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Workload	139
Abbildung 7.69 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Augenaktivität.....	140
Abbildung 7.70 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität.....	140
Abbildung 7.71 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Situationsbewusstsein.....	141
Abbildung 7.72 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Workload.....	141
Abbildung 7.73 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Augenaktivität.....	142
Abbildung 7.74 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität.....	142
Abbildung 7.75 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Situationsbewusstsein	143
Abbildung 7.76 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Workload	143
Abbildung A.1 Orientierungshilfe zur Interpretation der MTMM-Matrizen.....	199

TABELLENVERZEICHNIS

Tabelle 3.1 Methoden-Inventar zur Erhebung der Qualitätskriterien Situationsbewusstsein, Workload und Gebrauchstauglichkeit	41
Tabelle 4.1 Gegenüberstellung der Vor- und Nachteile experimenteller Vergleiche in der Simulationsumgebung	46
Tabelle 5.1 Methoden-Inventar zur Erhebung der Qualitätskriterien in der Machbarkeitsstudie	54
Tabelle 5.2 Methoden-Inventar zur Erhebung der Qualitätskriterien in der Re-Design-Studie	56
Tabelle 5.3 Gegenüberstellung der PSSUQ-Mediane der Machbarkeitsstudie (MS) und der Re-Design-Studie (RD)	76
Tabelle 5.4 Methoden-Inventar zur Erhebung der Qualitätskriterien in einer Studie zur Evaluierung eines räumlich-auditiven Warnsignals	83
Tabelle 5.5 Versuchsplan nach Rabl, Neujahr & Zimmer, 2016 zur Evaluierung eines räumlich-auditiven Warnsignals bei unterschiedlicher Belastung	83
Tabelle 5.6 Methoden-Inventar zur Erhebung der Qualitätskriterien in einer Studie zur Evaluierung der Bedienung eines Sensors	86
Tabelle 7.1 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Erfassung des Situationsbewusstseins	100
Tabelle 7.2 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Erfassung des Situationsbewusstseins	100
Tabelle 7.3 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Workload	101
Tabelle 7.4 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Situationsbewusstsein und Gebrauchstauglichkeit	102
Tabelle 7.5 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Workload	103
Tabelle 7.6 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Workload	103
Tabelle 7.7 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Situationsbewusstsein	104
Tabelle 7.8 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Workload und Gebrauchstauglichkeit	105
Tabelle 7.9 Prozentualer Anteil an Korrelationsstärken in den MMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit	106
Tabelle 7.10 Prozentualer Anteil an Korrelationsstärken in den HMMT-Matrizen zur Beschreibung der Gebrauchstauglichkeit	107

Tabellenverzeichnis

Tabelle 7.11 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Situationsbewusstsein	108
Tabelle 7.12 Prozentualer Anteil an Korrelationsstärken in den HMHT-Matrizen zur Beschreibung der diskriminanten Validität in der Kombination von Gebrauchstauglichkeit und Workload	108
Tabelle 7.13 Übersicht der validierten Methoden des Methoden-Inventars	111
Tabelle 7.14 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des 3D SARTs.....	112
Tabelle 7.15 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des 3D SARTs.....	112
Tabelle 7.16 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Workload.....	113
Tabelle 7.17 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des 3D SARTs in Kombination mit Gebrauchstauglichkeit.....	113
Tabelle 7.18 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen des SAGATs.....	113
Tabelle 7.19 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des SAGATs .	113
Tabelle 7.20 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Workload.....	114
Tabelle 7.21 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des SAGATs in Kombination mit Gebrauchstauglichkeit	114
Tabelle 7.22 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Sortieraufgabe	115
Tabelle 7.23 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Gefahren-Detektierung.....	116
Tabelle 7.24 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit	117
Tabelle 7.25 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit	117
Tabelle 7.26 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Workload.....	118
Tabelle 7.27 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit	118
Tabelle 7.28 Prozentuale Verteilung der Korrelationsstärken in den MMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit	118
Tabelle 7.29 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit	118
Tabelle 7.30 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Workload .	119
Tabelle 7.31 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung aufgabenbasierte Aufmerksamkeit in Kombination mit Gebrauchstauglichkeit	119

Tabellenverzeichnis

Tabelle 7.32	Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen des NASA TLX..	120
Tabelle 7.33	Prozentuale Angabe der Korrelationsstärke in den HMMT-Matrizen des NASA TLX.....	120
Tabelle 7.34	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Situationsbewusstsein.....	122
Tabelle 7.35	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des NASA TLX in Kombination mit Gebrauchstauglichkeit.....	122
Tabelle 7.36	Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen der TDRT	122
Tabelle 7.37	Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen der TDRT	122
Tabelle 7.38	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Situationsbewusstsein	123
Tabelle 7.39	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der TDRT in Kombination mit Gebrauchstauglichkeit	123
Tabelle 7.40	Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung	124
Tabelle 7.41	Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung	124
Tabelle 7.42	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Situationsbewusstsein	126
Tabelle 7.43	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Pupillenerweiterung in Kombination mit Gebrauchstauglichkeit	126
Tabelle 7.44	Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung PERCLOS.....	126
Tabelle 7.45	Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung PERCLOS.....	126
Tabelle 7.46	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Situationsbewusstsein.....	127
Tabelle 7.47	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung PERCLOS in Kombination mit Gebrauchstauglichkeit.....	127
Tabelle 7.48	Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung.....	128
Tabelle 7.49	Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung.....	128
Tabelle 7.50	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Situationsbewusstsein	129
Tabelle 7.51	Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking-Messung Aufmerksamkeitsfixierung in Kombination mit Gebrauchstauglichkeit	129
Tabelle 7.52	Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Zeitleistung	130
Tabelle 7.53	Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Zeitleistung	130

Tabellenverzeichnis

Tabelle 7.54 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Zeitleistung in Kombination mit Situationsbewusstsein.....	132
Tabelle 7.55 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Zeitleistung in Kombination mit Workload.....	132
Tabelle 7.56 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Geschwindigkeitsleistung in der 3DA-Studie.....	132
Tabelle 7.57 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Winkelgenauigkeit (Notch) in der 3DA-Studie.....	133
Tabelle 7.58 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Steuereingaben in der 3DA-Studie.....	134
Tabelle 7.59 Gegenüberstellung der Multi-Method/ Multi-Trait-Matrizen zur Erklärung der Genauigkeitsmessung Azimut in der 3DA-Studie.....	135
Tabelle 7.60 Prozentuale Verteilung der Korrelationsstärken in den MMMT-Matrizen des PSSUQ..	136
Tabelle 7.61 Prozentuale Verteilung der Korrelationsstärken in den HMMT-Matrizen des PSSUQ...	136
Tabelle 7.62 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Situationsbewusstsein.....	137
Tabelle 7.63 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen des PSSUQ in Kombination mit Workload.....	137
Tabelle 7.64 Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Eye-Tracking Zeitleistung.....	138
Tabelle 7.65 Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Eye-Tracking Zeitleistung.....	138
Tabelle 7.66 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Situationsbewusstsein.....	139
Tabelle 7.67 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Eye-Tracking Zeitleistung in Kombination mit Workload.....	139
Tabelle 7.68 Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der Augenaktivität.....	139
Tabelle 7.69 Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der Augenaktivität.....	139
Tabelle 7.70 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Situationsbewusstsein.....	141
Tabelle 7.71 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der Augenaktivität in Kombination mit Workload.....	141
Tabelle 7.72 Prozentuale Angabe der Korrelationsstärken in den MMMT-Matrizen der aufgabenbasierten Augenaktivität.....	141
Tabelle 7.73 Prozentuale Angabe der Korrelationsstärken in den HMMT-Matrizen der aufgabenbasierten Augenaktivität.....	141
Tabelle 7.74 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Situationsbewusstsein.....	142
Tabelle 7.75 Prozentuale Verteilung der Korrelationsstärken in den HMHT-Matrizen der aufgabenbasierten Augenaktivität in Kombination mit Workload.....	142

Tabellenverzeichnis

Tabelle A.1 Übersicht der verwendeten Submatrizen und deren Positionierung in der MTMM-Matrix.....	200
Tabelle A.2 Farbcodierung der Korrelationsstärken in den MTMM-Matrizen.....	200
Tabelle A.3 Farbcodierung der Signifikanzen in den MTMM-Matrizen	200
Tabelle B.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Machbarkeitsstudie	201
Tabelle B.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Machbarkeitsstudie	201
Tabelle B.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Machbarkeitsstudie	202
Tabelle B.4 MTMM-Matrix der Machbarkeitsstudie.....	203
Tabelle B.5 Signifikanz-Matrix der Machbarkeitsstudie.....	204
Tabelle C.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Re-Design-Studie.....	205
Tabelle C.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Re-Design-Studie	206
Tabelle C.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Re-Design-Studie.....	207
Tabelle C.4 MTMM-Matrix der Re-Design-Studie	208
Tabelle C.5 Signifikanz-Matrix der Re-Design-Studie	209
Tabelle D.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der 3DA-Studie	210
Tabelle D.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der 3DA-Studie....	211
Tabelle D.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der 3DA-Studie	212
Tabelle D.4 MTMM-Matrix der 3DA-Untersuchung.....	213
Tabelle D.5 Signifikanz-Matrix der 3DA-Untersuchung.....	214
Tabelle E.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Studie zur Bedienung eines Sensors	215
Tabelle E.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Studie zur Bedienung eines Sensors	216
Tabelle E.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Studie zur Bedienung eines Sensors	217
Tabelle E.4 MTMM-Matrix der Sensor-Studie.....	218
Tabelle E.5 Signifikanz-Matrix der Sensor-Studie.....	218

A ORIENTIERUNGSHILFE ZUR INTERPRETATION DER MULTI-TRAIT/MULTI-METHOD-MATRIZEN

In den Anhängen B, C, D und E sind jeweils eine tabellarische Übersicht der verwendeten Methoden und Variablen, die Multi-Trait/ Multi-Method (MTMM)-Matrizen und deren Submatrizen sowie die zugehörigen Signifikanzen und deren Matrizen enthalten. Als Submatrizen sind hier, wie in Kapitel 6 beschrieben, die MonoMethod-MonoTrait (MMMT)-Matrizen, die HeteroMethod-MonoTrait (HMMT)-Matrizen und die HeteroMethod-HeteroTrait (HMHT)-Matrizen beschrieben, die jeweils eine MTMM-Matrix beziehungsweise eine Evaluierungsmatrix bilden.

Abbildung A.1 dient hierzu als Orientierungshilfe zur Interpretation der MTMM-Matrizen für die durchgeführten Evaluierungen. Die jeweiligen Bezeichnungen der Submatrizen sowie deren betroffenen Qualitätskriterien sind in Tabelle A.1 mit Bezug auf die jeweilige Nummerierung zusammengefasst und die verwendete Farbkodierung für die Korrelationsstärken und deren Signifikanzen werden in Tabelle A.2 und Tabelle A.3 dargestellt. Die Farbkodierung kann zur Interpretation der MTMM-Matrizen in den Anhängen B, C, D und E genutzt werden, weil eine Darstellung der Daten in ausreichender Schriftgröße auf Grund deren Größe nicht möglich ist. Da jedoch die Auswertung auf Grundlage der vorgegebenen, farbkodierten Datenbereiche vorgenommen worden ist, ist eine exakte Darstellung der Daten nachrangig, weil kein zusätzlicher Informationsgewinn erfolgt.

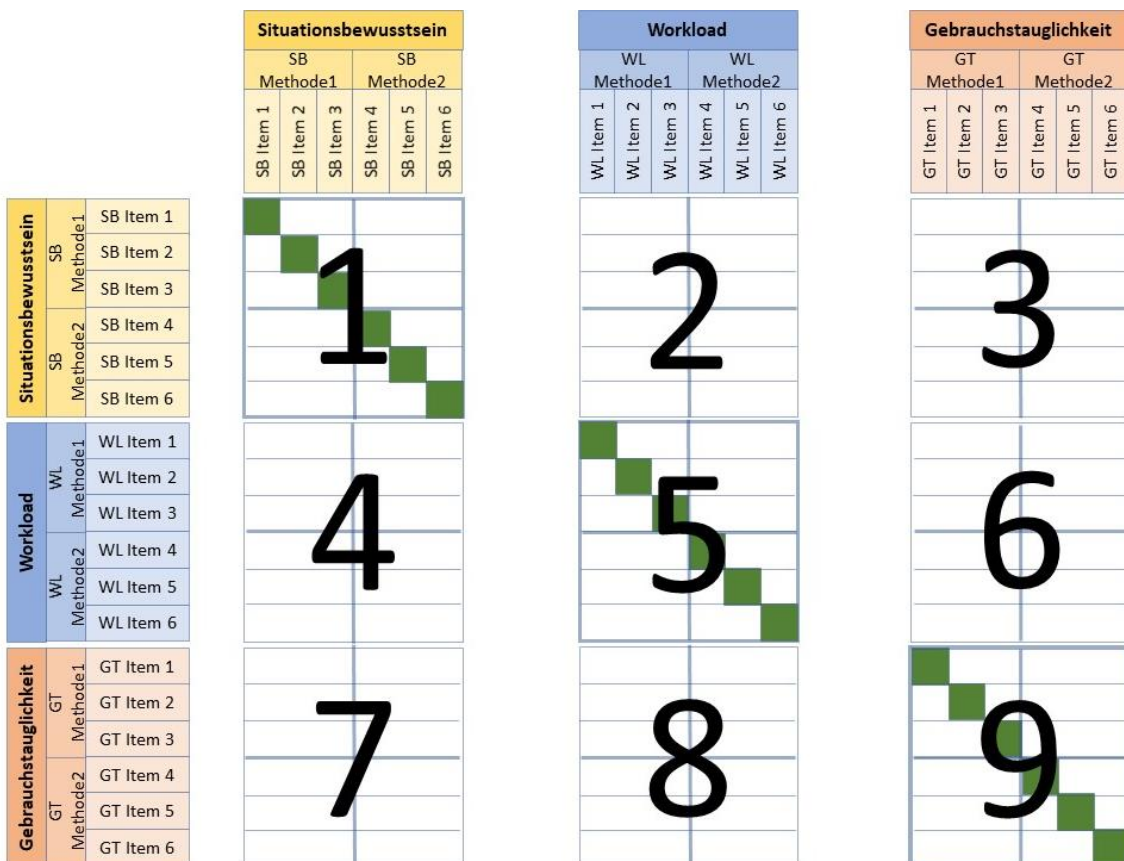


Abbildung A.1 Orientierungshilfe zur Interpretation der MTMM-Matrizen

A Orientierungshilfe zur Interpretation der Multi-Trait/Multi-Method-Matrizen

Tabelle A.1 Übersicht der verwendeten Submatrizen und deren Positionierung in der MTMM-Matrix

Submatrix	Bezeichnung	Betroffene Konstrukte
1	MMMT-Matrizen HMMT-Matrizen	Situationsbewusstsein
2	HMHT-Matrizen	Situationsbewusstsein – Workload
3	HMHT-Matrizen	Situationsbewusstsein – Gebrauchstauglichkeit
4	HMHT-Matrizen	Workload – Situationsbewusstsein
5	MMMT-Matrizen HMMT-Matrizen	Workload
6	HMHT-Matrizen	Workload – Gebrauchstauglichkeit
7	HMHT-Matrizen	Gebrauchstauglichkeit – Situationsbewusstsein
8	HMHT-Matrizen	Gebrauchstauglichkeit – Workload
9	MMMT-Matrizen HMMT-Matrizen	Gebrauchstauglichkeit

Tabelle A.2 Farbcodierung der Korrelationsstärken in den MTMM-Matrizen

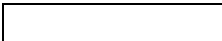








Farbkodierung	Spearman's ρ
	$ \rho \approx 0$
	$0.2 < \rho \leq 0.4$
	$0.4 < \rho \leq 0.6$
	$ \rho > 0.6$
	$ \rho = 1$

Tabelle A.3 Farbcodierung der Signifikanzen in den MTMM-Matrizen

Farbkodierung	p-Value
	Diagonale der Eigenkorrelationen
	$p > .05$
	$p \leq .05$
	$p \leq .01$

B MULTI-TRAIT/MULTI-METHOD-MATRIX DER MACHBARKEITSSTUDIE

Tabelle B.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Machbarkeitsstudie

Situationsbewusstsein	EAK	Eye-Tracking Aufmerksamkeit	NrG	Anzahl der Blicke
			AR	Prozentuale Aufmerksamkeit auf ein AOI
			NrF	Anzahl der Fixation
			FD	Fixationsdauer
	AAK	Eye-Tracking aufgabenbasierte Aufmerksamkeit	INrG	Anzahl der Blicke während der Aufgabe ‚Eingabe‘
			INF	Anzahl der Fixationen während der Aufgabe ‚Eingabe‘
			IFD	Fixationsdauer während der Aufgabe ‚Eingabe‘

Tabelle B.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Machbarkeitsstudie

Workload	TLX	NASA TLX	MD	Mentale Anforderungen
			PD	Physische Anforderungen
			TD	Zeitliche Anforderungen
			P	Leistung
			E	Ressourceneinsatz
			F	Frustrationslevel
			OVR	Workload Gesamtwert (raw)
	PD	Pupillen Dilatation	PA	Pupillenfläche
			PH	Pupillenhöhe
			PW	Pupillenweite
	PCL	PERCLOS	PCa	Anteil PERCLOS Lidschluss
			IPC	Anteil PERCLOS Lidschluss während der Aufgabe ‚Eingabe‘
	AF	Aufmerksamkeits- fixierung	MGD	Maximale Blickdauer
G2s			Anzahl an Blicken, länger als 2 Sekunden	

B Multi-Trait/Multi-Method-Matrix der Machbarkeitsstudie

Tabelle B.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Machbarkeitsstudie

Gebrauchstauglichkeit	ZT	Leistung Zeit	TTm ITm	Mittlere Bearbeitungsdauer der Aufgaben Mittlere Bearbeitungsdauer der Aufgabe ‚Eingabe‘
	EZT	Eye-Tracking Leistung Zeit	TGT mG D GRs IGD	Gesamtblickdauer Mittlere Blickdauer Blicke pro Sekunde Mittlere Blickdauer während der Aufgabe ‚Eingabe‘
	EAT	Eye-Tracking Augenaktivität	%TT EAY HEY VEY NrS SL	Prozentualer Anteil an Transitionszeiten Augenaktivität pro Sekunde Horizontale Augenaktivität Vertikale Augenaktivität Anzahl an Sakkaden Länge der Sakkaden
	AAT	Eye-Tracking aufgabenba- sierende Augen- aktivität	INrS ISL	Anzahl an Sakkaden während der Aufgabe ‚Eingabe‘ Länge der Sakkaden während der Aufgabe ‚Eingabe‘
	PSQ	PSSUQ	P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15 P16 SU INQ ITQ OV	Zufriedenheit mit der Bedienung Einfachheit der Bedienung Geschwindigkeit der Aufgabenerfüllung Bequeme Bedienung Einfachheit des Erlernens der Bedienung Geschwindigkeit, mit der ein produktiver Einsatz mög- lich ist Qualität der Fehlermeldungen Eingabefehler konnten rückgängig gemacht werden Klare Dokumentation Verfügbarkeit der Informationen Hilfreiche Informationen zur Aufgabenerfüllung Organisation der Informationen Ansprechbarkeit der Bedienoberfläche Nutzerfreundlichkeit der Bedienoberfläche Vollständigkeit des Produktes Nutzerzufriedenheit mit dem Produkt Nützlichkeit des Systems Informationsqualität Qualität der Schnittstelle Gesamtwert Gebrauchstauglichkeit

C MULTI-TRAIT/MULTI-METHOD-MATRIX DER RE-DESIGN-STUDIE

Tabelle C.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Re-Design-Studie

Situationsbewusstsein	SRT	3D SART	DAR SAR US SA	Beanspruchung von Aufmerksamkeitsressourcen Unterstützung von Aufmerksamkeitsressourcen Situationsverständnis Situationsbewusstsein gesamt
	SGT	SAGAT	ITM %RT %PT %PS SUPS fWP HWP MWP CWP FAWP HSUW	SAGAT-Items zu Frequenzwechsel, Höhenänderungen und QNH Prozentual korrekt erinnerte Routenabschnitte Prozentual korrekt erinnerte Navigationspunkte ohne Namen Prozentual korrekt erinnerte Position der Punkte mit Namen Sicherheitsurteil bezüglich der zugewiesenen Positionen Anteil falsch erinnertes Navigationspunkte Hit-Rate der Navigationspunkte Miss-Rate der Navigationspunkte Rate der korrekt zurückgewiesenen Navigationspunkte Rate der fälschlich identifizierten Navigationspunkte Sicherheitsurteil der Hits bezogen auf alle Navigationspunkte
	SA	Sortieraufgabe	SCN DIE LRN	Szenarioabhängige Sortierung Ideale Sortierung Lernabhängige Sortierung
	EAK	Eye-Tracking Aufmerksamkeit	NrG AR NrFl NrFr mFDl mFDr	Anzahl der Blicke Prozentuale Aufmerksamkeitsverteilung auf ein AOI Anzahl an Fixationen (links) Anzahl an Fixationen (rechts) Mittlere Fixationsdauer (links) Mittlere Fixationsdauer (rechts)
	AAK	Eye-Tracking aufgaben- basierte Aufmerksamkeit	INrG IFDI IFDr INFl INFr RNrG RFDI RFDr RNFl RNFr	Anzahl der Blicke während der Aufgabe ‚Eingabe‘ Mittlere Fixationsdauer (links) während der Aufgabe ‚Eingabe‘ Mittlere Fixationsdauer (rechts) während der Aufgabe ‚Eingabe‘ Anzahl an Fixationen (links) während der Aufgabe ‚Eingabe‘ Anzahl an Fixationen (rechts) während der Aufgabe ‚Eingabe‘ Anzahl der Blicke während der Aufgabe ‚Umleitung‘ Mittlere Fixationsdauer (links) während der Aufgabe ‚Umleitung‘ Mittlere Fixationsdauer (rechts) während der Aufgabe ‚Umleitung‘ Anzahl an Fixationen (links) während der Aufgabe ‚Umleitung‘ Anzahl an Fixationen (rechts) während der Aufgabe ‚Umleitung‘

Tabelle C.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Re-Design-Studie

Workload	TLX	NASA TLX	MD PD TD P E F OVR	Mentale Anforderungen Physische Anforderungen Zeitliche Anforderungen Leistung Ressourcen-Einsatz Frustrationslevel Workload Gesamtwert (raw)
	DRT	TDRT	TP MR HR HRKT mRKT	Anzahl erfolgreicher Trials Relative Miss-Rate Relative Hit-Rate Hit Reaktionszeit Reaktionszeit
	PD	Pupillen Dilatation	IPA rPA IPH rPH IPW rPW	Pupillenfläche (links) Pupillenfläche (rechts) Pupillenhöhe (links) Pupillenhöhe (rechts) Pupillenweite (links) Pupillenweite (rechts)
	PCL	PERCLOS	PCI PCr PCa	Anteil PERCLOS Lidschluss (links) Anteil PERCLOS Lidschluss (rechts) Durchschnittl. Anteil PERCLOS Lidschluss
	AF	Aufmerksamkeits- fixierung	MGD G2s	Maximale Blickdauer Anzahl an Blicken, länger als 2 Sekunden

C Multi-Trait/Multi-Method-Matrix der Re-Design-Studie

Tabelle C.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Re-Design-Studie

Gebrauchstauglichkeit	ZT	Leistung Zeit	ITM	Eingabezeit
	EZT	Eye-Tracking Leistung Zeit	TGT mGD GRs Idur IGD Rdur RGD	Gesamtblickdauer Mittlere Blickdauer Blicke pro Sekunde Dauer während der Aufgabe ‚Eingabe‘ Mittlere Blickdauer während der Aufgabe ‚Eingabe‘ Dauer während der Aufgabe ‚Umleitung‘ Mittlere Blickdauer während der Aufgabe ‚Umleitung‘
	EAT	Eye-Tracking Augenaktivität	%TT HEY VEY NrSl NrSr SDI SDr SAI SAr	Prozentualer Anteil an Transitionszeiten Horizontale Augenaktivität Vertikale Augenaktivität Anzahl an Sakkaden (links) Anzahl an Sakkaden (rechts) Mittlere Sakkaden-Dauer (links) Mittlere Sakkaden-Dauer (rechts) Sakkaden-Winkel (links) Sakkaden-Winkel (rechts)
	AAT	Eye-Tracking aufgabenbasierte Augenaktivität	ISDI ISDr INSI INSr RNSI RNSr RSDI RSDr	Mittlere Sakkaden-Dauer (links) während der Aufgabe ‚Eingabe‘ Mittlere Sakkaden-Dauer (rechts) während der Aufgabe ‚Eingabe‘ Anzahl an Sakkaden (links) während der Aufgabe ‚Eingabe‘ Anzahl an Sakkaden (rechts) während der Aufgabe ‚Eingabe‘ Anzahl an Sakkaden (links) während der Aufgabe ‚Umleitung‘ Anzahl an Sakkaden (rechts) während der Aufgabe ‚Umleitung‘ Mittlere Sakkaden-Dauer (links) während der Aufgabe ‚Umleitung‘ Mittlere Sakkaden-Dauer (rechts) während der Aufgabe ‚Umleitung‘
	PSQ	PSSUQ	P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15 P16 SU INQ ITQ OV	Zufriedenheit mit der Bedienung Einfachheit der Bedienung Geschwindigkeit der Aufgabenerfüllung Bequeme Bedienung Einfachheit des Erlernens der Bedienung Geschwindigkeit, mit der ein produktiver Einsatz möglich ist Qualität der Fehlermeldungen Eingabefehler konnten rückgängig gemacht werden Klare Dokumentation Verfügbarkeit der Informationen Hilfreiche Informationen zur Aufgabenerfüllung Organisation der Informationen Ansprechbarkeit der Bedienoberfläche Nutzerfreundlichkeit der Bedienoberfläche Vollständigkeit des Produktes Nutzerzufriedenheit mit dem Produkt Nützlichkeit des Systems Informationsqualität Qualität der Schnittstelle Gesamtwert Gebrauchstauglichkeit

D MULTI-TRAIT/MULTI-METHOD-MATRIX DER 3DA-UNTERSUCHUNG

Tabelle D.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der 3DA-Studie

Situationsbewusstsein	SRT	3D SART	DAR US SAR SA	Beanspruchung von Aufmerksamkeitsressourcen Situationsverständnis Unterstützung von Aufmerksamkeitsressourcen Situationsbewusstsein gesamt
	SGT	SAGAT	SRKT EstD EinR DinR RinR ECSp RCSp DCSp EPer RPer DPer EHd RAZ DHd	Reaktionsgeschwindigkeit auf SAGAT-Fragen Einschätzungsabweichung über alle SAGAT-Fragen Einschätzung der Initialreaktion durch den Probanden Abweichung der Initialreaktion Reale Initialreaktion Zeiteinschätzung, bis Zielfluglage erreicht wurde Benötigte Zeit, bis Zielfluglage erreicht wurde Abweichung der Zeiteinschätzung, bis Zielfluglage erreicht wurde Einschätzung der eigenen Leistung Reale Bewertung der Leistung Abweichung der Leistungseinschätzung Richtungseinschätzung der Gefahrenquelle Reales Auftreten der Gefahrenquelle Abweichung zum realen Auftreten der Gefahrenquelle
	GD	Gefahren-Detektierung	MmDN RVL RP DRT	Gemittelte Anzahl an erkannten Gefahren Bewertung durch den Versuchsleiter Bewertung durch den Probanden Abweichung der Bewertung
	EAK	Eye-Tracking Aufmerksamkeit	NrG AR NrFl NrFr mFDl mFDr	Anzahl der Blicke Prozentuale Aufmerksamkeitsverteilung auf ein AOI Anzahl der Fixationen (links) Anzahl der Fixationen (rechts) Mittlere Fixationsdauer (links) Mittlere Fixationsdauer (rechts)

D Multi-Trait/Multi-Method-Matrix der 3DA-Untersuchung

Tabelle D.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der 3DA-Studie

Workload	TLX	NASA TLX	MD PD TD P E F OVR	Mentale Anforderungen Physische Anforderungen Zeitliche Anforderungen Leistung Ressourcen-Einsatz Frustrationslevel Workload Gesamtwert (raw)
	DRT	TDRT	TRL MR HR HRKT	Anzahl erfolgreicher Trials Relative Miss-Rate Relative Hit-Rate Hit Reaktionszeit
	PD	Pupillen Dilatation	IPA rPA IPH rPH IPW rPW	Pupillenfläche (links) Pupillenfläche (rechts) Pupillenhöhe (links) Pupillenhöhe (rechts) Pupillenweite (links) Pupillenweite (rechts)
	PCL	PERCLOS	PCI PCr PCa	Anteil PERCLOS Lidschluss (links) Anteil PERCLOS Lidschluss (rechts) Durchschnittl. Anteil PERCLOS Lidschluss
	AF	Aufmerksamkeitsfixierung	MGD G2s	Maximale Blickdauer Anzahl an Blicken, länger als 2 Sekunden

Tabelle D.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der 3DA-Studie

Gebrauchstauglichkeit	GKT	Leistung Geschwindigkeit	mSp vSp	Im Mittel geflogene Geschwindigkeit Varianz der Geschwindigkeit
	NTC	Leistung Notch	NDO NDR PND NAD NDE NDVm	Benötigter Drehwinkel zur Gefahrenquelle Geflogener Drehwinkel zur Gefahrenquelle Abweichung des Drehwinkel Abweichungen unter 10° Abweichungen unter 5° Geringste Abweichung des Drehwinkel
	STE	Leistung Steuereingabe	azSt alSt	Anzahl der Steuerausschläge +/- 90° vom Nullpunkt Anzahl der Steuerausschläge +/- zur Ausgangshöhe
	AZ	Leistung Azimut	aR10 aR5 plAz rIAz rIDe	Anzahl der Eintritte kleiner oder gleich 10° Anzahl der Eintritte kleiner oder gleich 5° Geplant zu fliegende Leistung im Azimut Real geflogene Leistung im Azimut Abweichung zur geplanten Leistung im Azimut
	ZT	Leistung Zeit	TM TN10 TN5 aT10 aT5 aTO eT25 eTO tT	Reaktionszeit bis Beginn des Flugmanövers Zeit bis Drehwinkel mit 10° Abweichung erreicht wurde Zeit bis Drehwinkel mit 5° Abweichung erreicht wurde Zeit innerhalb von 10° Abweichung Zeit innerhalb von 5° Abweichung Zeit außerhalb der Qualitätskriterien Zeit innerhalb des Höhenbandes von 250 m Zeit außerhalb des Höhenbandes Dauer eines Trials
	EZT	Eye-Tracking Leistung Zeit	TGT mGD GRs	Gesamtblickdauer Mittlere Blickdauer Blicke pro Sekunde
	EAT	Eye-Tracking Augenaktivität	%TT HEY VEY NrSl NrSr SDI SDr SAI SAr	Prozentualer Anteil an Transitionszeiten Horizontale Augenaktivität Vertikale Augenaktivität Anzahl an Sakkaden (links) Anzahl an Sakkaden (rechts) Mittlere Sakkaden-Dauer (links) Mittlere Sakkaden-Dauer (rechts) Sakkaden-Winkel (links) Sakkaden-Winkel (rechts)

E MULTI-TRAIT/MULTI-METHOD-MATRIX DER SENSOR-STUDIE

Tabelle E.1 Methoden- und Variablen-Übersicht zur Erhebung von Situationsbewusstsein in der Studie zur Bedienung eines Sensors

Situationsbewusstsein	SRT	3D SART	DAR SAR US SA	Beanspruchung von Aufmerksamkeitsressourcen Unterstützung von Aufmerksamkeitsressourcen Situationsverständnis Situationsbewusstsein gesamt
	EAK	Eye-Tracking Aufmerksamkeit	NrG AR NrFl NrFr mFDI mFDr	Anzahl der Blicke Prozentuale Aufmerksamkeitsverteilung auf ein AOI Anzahl an Fixationen (links) Anzahl an Fixationen (rechts) Mittlere Fixationsdauer (links) Mittlere Fixationsdauer (rechts)
	AAK	Eye-Tracking aufgabenbasierte Aufmerksamkeit	ING IFDI IFDr INFI INFr TNG TFDI TFDr TNFI TNFr STNG	Anzahl der Blicke während der Aufgabe ‚Sensitivitätseinstellung‘ Mittlere Fixationsdauer (links) während der Aufgabe ‚Sensitivitätseinstellung‘ Mittlere Fixationsdauer (rechts) während der Aufgabe ‚Sensitivitätseinstellung‘ Anzahl an Fixationen (links) während der Aufgabe ‚Sensitivitätseinstellung‘ Anzahl an Fixationen (rechts) während der Aufgabe ‚Sensitivitätseinstellung‘ Anzahl der Blicke während der Aufgabe ‚Genauigkeitseinstellung‘ Mittlere Fixationsdauer (links) während der Aufgabe ‚Genauigkeitseinstellung‘ Mittlere Fixationsdauer (rechts) während der Aufgabe ‚Genauigkeitseinstellung‘ Anzahl an Fixationen (links) während der Aufgabe ‚Genauigkeitseinstellung‘ Anzahl an Fixationen (rechts) während der Aufgabe ‚Genauigkeitseinstellung‘ Anzahl der Blicke während der Zweitaufgabe

E Multi-Trait/Multi-Method-Matrix der Sensor-Studie

Tabelle E.2 Methoden- und Variablen-Übersicht zur Erhebung von Workload in der Studie zur Bedienung eines Sensors

Workload	TLX	NASA TLX	MD PD TD P E F OVR	Mentale Anforderungen Physische Anforderungen Zeitliche Anforderungen Leistung Ressourcen-Einsatz Frustrationslevel Workload Gesamtwert (raw)
	PD	Pupillen Dilatation	IPA rPA IPH rPH IPW rPW	Pupillenfläche (links) Pupillenfläche (rechts) Pupillenhöhe (links) Pupillenhöhe (rechts) Pupillenweite (links) Pupillenweite (rechts)
	PCL	PERCLOS	PCI PCa PCr	Anteil PERCLOS Lidschluss (links) Durchschnittl. Anteil PERCLOS Lidschluss Anteil PERCLOS Lidschluss (rechts)
	AF	Aufmerksamkeits- fixierung	MGD G2s	Maximale Blickdauer Anzahl an Blicken, länger als 2 Sekunden

Tabelle E.3 Methoden- und Variablen-Übersicht zur Erhebung von Gebrauchstauglichkeit in der Studie zur Bedienung eines Sensors

Gebrauchstauglichkeit	ZT	Leistung Zeit	TRKT	Bearbeitungszeit während der Aufgabe ‚Genauigkeitseinstellung‘
			IRKT	Bearbeitungszeit während der Aufgabe ‚Sensitivitätseinstellung‘
	EZT	Eye-Tracking Leistung Zeit	TGT	Gesamtblickdauer
			mGD	Mittlere Blickdauer
			GRs	Blicke pro Sekunde
			DNs	Dauer
			ImGD	Mittlere Blickdauer während der Aufgabe ‚Sensitivitätseinstellung‘
			TmGD	Mittlere Blickdauer während der Aufgabe ‚Genauigkeitseinstellung‘
			STGD	Mittlere Blickdauer während der Zweitaufgabe
	ETA	Eye-Tracking Augenaktivität	%TT	Prozentualer Anteil an Transitionszeiten
			HEY	Horizontale Augenaktivität
			VEY	Vertikale Augenaktivität
			NrSl	Anzahl an Sakkaden (links)
			NrSr	Anzahl an Sakkaden (rechts)
			SDI	Mittlere Sakkaden-Dauer (links)
			SDr	Mittlere Sakkaden-Dauer (rechts)
			SAI	Sakkaden-Winkel (links)
			SAr	Sakkaden-Winkel (rechts)
	AAT	Eye-Tracking aufgabenbasierte Augenaktivität	INSI	Anzahl an Sakkaden (links) während der Aufgabe ‚Sensitivitätseinstellung‘
			INSr	Anzahl an Sakkaden (rechts) während der Aufgabe ‚Sensitivitätseinstellung‘
			ISDI	Mittlere Sakkaden-Dauer (links) während der Aufgabe ‚Sensitivitätseinstellung‘
			ISDr	Mittlere Sakkaden-Dauer (rechts) während der Aufgabe ‚Sensitivitätseinstellung‘
			TNSI	Anzahl an Sakkaden (links) während der Aufgabe ‚Genauigkeitseinstellung‘
			TNSr	Anzahl an Sakkaden (rechts) während der Aufgabe ‚Genauigkeitseinstellung‘
			TSDI	Mittlere Sakkaden-Dauer (links) während der Aufgabe ‚Genauigkeitseinstellung‘
			TSDr	Mittlere Sakkaden-Dauer (rechts) während der Aufgabe ‚Genauigkeitseinstellung‘

